How unconscious biases can skew forensic science p. 686

Hydrodynamics emerges in spin chains pp. 699, 716, & 720

Global shifts in coastal wetlands p. 744

Science

$15
13 MAY 2022
science.org

AAAS

HUMAN CELL ATLAS

Mapping cell identity across tissues and organs pp. 695 & 711–713

2023 AWARDS & PRIZES

AAAS DAVID AND BETTY HAMBURG AWARD FOR SCIENCE DIPLOMACY

SUBMIT A NOMINATION TODAY
Deadline June 30, 2022

The **AAAS David and Betty Hamburg Award for Science Diplomacy** recognizes an individual or a limited number of individuals working together in the scientific and engineering or foreign affairs communities who are making an outstanding contribution to furthering science diplomacy.

Over the past 30 years, AAAS has honored an international cadre of science luminaries for their contributions to international scientific cooperation and science diplomacy. **In 2021, the award was renamed after David and Betty Hamburg** to recognize their unparalleled commitment to the significant role of science diplomacy to advance science, human rights, peace, and cooperation.

- The award is open to all regardless of nationality or citizenship.
- We accept self-nominations.
- Nominees must be living at the time of their nomination.

To learn more, visit

aaas.org/awards/science-diplomacy/about

AAAS gratefully acknowledges Carnegie Corporation of New York for their generous support to launch the AAAS David and Betty Hamburg Award for Science Diplomacy and the individuals and foundations whose contributions have begun an endowment that will allow us to sustain it in perpetuity.

PHOTO CREDIT: DIRCO

2016 WINNER

Minister Naledi Pandor, Ph.D.
South African Minister of International Relations and Cooperation

Recognized for using science and technology to support development in South Africa and sub-Saharan Africa.

Winners will receive:

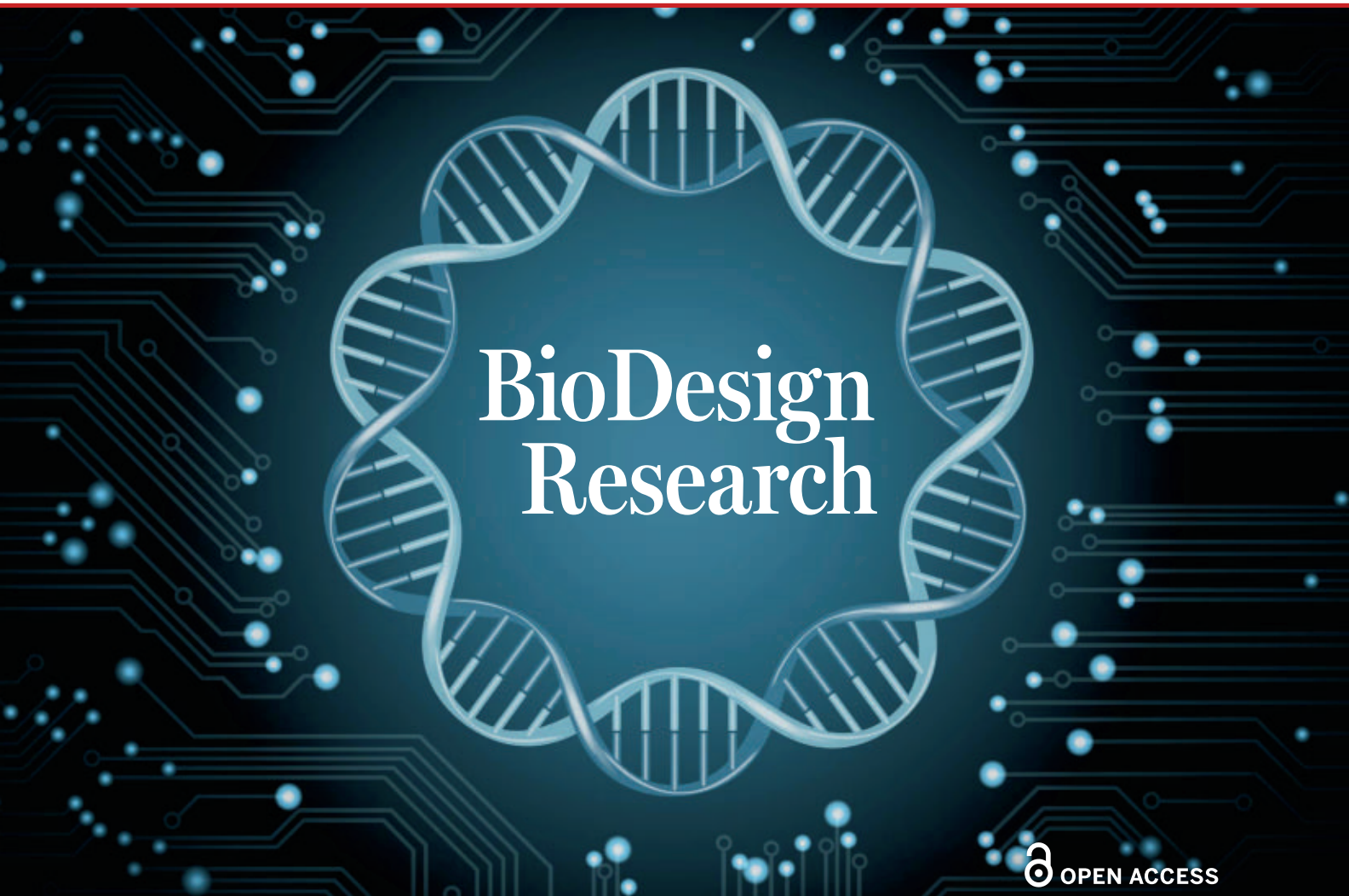
- Monetary prize of $10,000
- Commemorative plaque
- Worldwide promotion of their achievements through AAAS communication channels, including AAAS publications, member news, website, and social media
- Complimentary registration to the 2023 AAAS Annual Meeting
- Reimbursement for travel and hotel expenses to attend the AAAS Annual Meeting
- The opportunity to publish in *Science & Diplomacy*

AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS



OPEN ACCESS

BioDesign Research is a Science Partner Journal published in affiliation with **Nanjing Agricultural University (NAU)** and distributed by the **American Association for the Advancement of Science (AAAS).** *BioDesign Research* publishes high quality breakthrough research, reviews, editorials, and perspectives focusing on in silico biosystems design, genetic or epigenetic modifications, and genome writing or rewriting in any organism.

Submit your research to *BioDesign Research* today!

Learn more at **spj.sciencemag.org/bdr**

The Science Partner Journals (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the non-profit publisher of the *Science* family of journals. The SPJ program features high quality, online-only, editorially independent open-access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS expands its efforts to communicate science broadly and for the benefit of all people by providing a top-tier international research organization with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.

Learn more at **spj.sciencemag.org**

@BioDesignR @SPJournals

ARTICLE PROCESSING CHARGES WAIVED UNTIL 2022

The POSTECH team (left to right): Jun Sung Kim, Jonghwan Kim, Jong Kyu Kim, Moon-Ho Jo, Gil-Ho Lee

The future of new materials at POSTECH

Since the Pohang University of Science and Technology in South Korea, commonly called POSTECH, opened its doors in 1986, materials science and engineering research has been pivotal to its academic mission. POSTECH has eight institutes dedicated to materials research whose work regularly appears in top academic journals, and the faculty there are passionate about generating breakthroughs that transform how new technologies are developed.

The impact of POSTECH's materials research has the potential to transform everyday life, from the development of lighter, more energy-efficient and less expensive electronic devices to a reimagining of how we tackle human disease.

Materials scientist Moon-Ho Jo is the director of the Center for Epitaxial van der Waals Quantum Solids at POSTECH (*1*), a part of South Korea's Institute for Basic Science (IBS) that is dedicated to exploring new types of atomically thin van der Waals (vdW) materials.

At Jo's center, researchers will be looking to create novel classes of vdW materials to serve as platforms for next-generation quantum technologies. He explains that two-dimensional (2D) vdW materials are layered and can be stacked in different ways to form solids with new material properties, ranging from insulators and semiconductors to semimetals and superconductors. "I have set out to create heteroepitaxial vdW solids, where their crystal lattices and symmetries are artificially moulded with atomic precision, and can be translated for use in real-life devices," he says.

Putting a new spin on magnets

Condensed matter physicist Jun Sung Kim works in the Center for Artificial Low Dimensional Electron Systems, where he focuses on researching the novel properties of 2D vdW magnets.

Kim explains that these topological materials, whose electronic bands are described as "twisted," are currently one of the most interesting material classes. Topological materials have become important in spintronic applications, in which the electron's spin information is used rather than simply its charge information, explains Kim. It is expected that spintronics research will lead to the development of electronic devices that have increased data storage capabilities and faster processing speeds but consume less power—a change that will help reduce energy consumption worldwide and put the use of computing technologies on a more environmentally sustainable path. "We believe that novel properties in a new class of magnets could provide breakthroughs for developing faster, nonvolatile, and more energy-efficient information technology," he says.

Jo's research led to a recent paper published in *Nature* (*2*), which describes his team's discovery that the magnetic semiconductor $Mn_3Si_2Te_6$ exhibits unprecedentedly large angular magnetoresistance, often called colossal angular magnetoresistance (CAMR), due to its unique interplay between topological electronic structure and magnetism. "We hope that our series of discoveries on new topological

PHOTO: COURTESY OF POSTECH

van der Waals magnets will provide a material platform to investigate the rich topological aspects of two-dimensional magnets and their functionalities for next-generation electronics or spintronics," he says.

Optoelectronic devices

Researcher Jonghwan Kim leads a group (*3*) at POSTECH's Laboratory of Optoelectronics Materials and Devices. For him and his team, vdW materials provide exciting opportunities to explore new optical science and technology developments.

The crystal structure of these semiconductors is composed of a single or just a few atomic layers, making them the thinnest form of 2D materials in nature, he says. They have high light absorption and emission properties and can be flexibly and precisely stacked to create new materials.

Kim's research team recently developed a deep-ultraviolet (DUV) light-emitting diode (LED) using a new material. This DUV LED was designed to emit UV light at a short wavelength of 200 nm–280 nm, which could potentially be used to destroy disease in the human body. The team used hexagonal boron nitride (hBN), which is a vdW layered material like graphite, to fabricate an LED device that emits strong UV light.

"Throughout history, there have been many examples of new materials enabling groundbreaking change in real life," says Kim. In the future, Kim looks forward to the possibility of applying the technology in human health care to tackle dangerous viruses, such as COVID-19, by sterilizing these pathogens using a shortwave length of UV light that has a low level of skin penetration and is less harmful to the human body.

Next-generation electronics

Based in the Department of Materials Science and Engineering at POSTECH, Jong Kyu Kim leads a cutting-edge team focused on researching materials breakthroughs that could support the development of more electronic devices based on 2D materials. Kim is focusing on advancing our understanding of the growth mechanism of hBNs by metal-organic chemical vapor deposition (MOCVD), which he expects will be a step toward developing the production of high-quality hBN building blocks for use in a variety of silicon-based electronic devices (*4*). "Recently, an hBN layer was successfully grown on a 2-inch, high-electron-mobility transistor, and it performed well," he says.

Kim is confident that hBN could become a new material platform for supporting the development of smaller, faster, and cheaper electronic devices, and could help to overcome the current scaling limitation of silicon-based devices. "2D materials with a thickness of just a few nanometers could replace current electronic devices, while showing a higher performance and offering additional functionality."

Last year, Kim's team was also featured in a leading journal (*5*) when they proposed a way to efficiently produce hydrogen fuel via water electrolysis using cheap and readily available nickel as an electrocatalyst. "This research is significant in that it provides the scholarly foundation for high performance and commercialization of a sustainable hydrogen energy conversion system," explains Kim.

Diving deeper into physical properties

Researchers at POSTECH combine a passion to create real-life applications for their research with a fundamental desire to better understand how our natural world performs. Physicist Gil-Ho Lee (*6*) leads a quantum nanoelectronics lab at POSTECH. His research focuses on experimental studies on quantum devices based on vdW materials, including graphene, superconductors, and topological and magnetic materials.

> **"Researchers at POSTECH combine a passion to create real-life applications for their research with a fundamental desire to better understand how our natural world performs."**

Lee analyses the quantum behaviour of devices made of these materials and measures their electrical properties at low temperatures, to determine how these lessons could inform the development of computing or sensors at the nano level. He says: "As the field of electromagnetism in physics opened up the field of electrical engineering, so I hope quantum mechanics will open up the field of so-called quantum engineering."

Recently, Lee's team, together with collaborators, developed quantum sensors for microwave and infrared lights. This work, in which they proposed using graphene-based quantum sensors to exploit graphene's minute electronic heat capacity, was published in leading international journals (*7*, *8*). "This research is a good example of how a fundamental understanding of quantum materials can be used for sensor applications," explains Lee.

Together with his colleagues across POSTECH, Lee is looking for answers that could help researchers understand the quantum physics of nanodevices and how these solutions could support us in meeting some of the greatest challenges facing society.

References

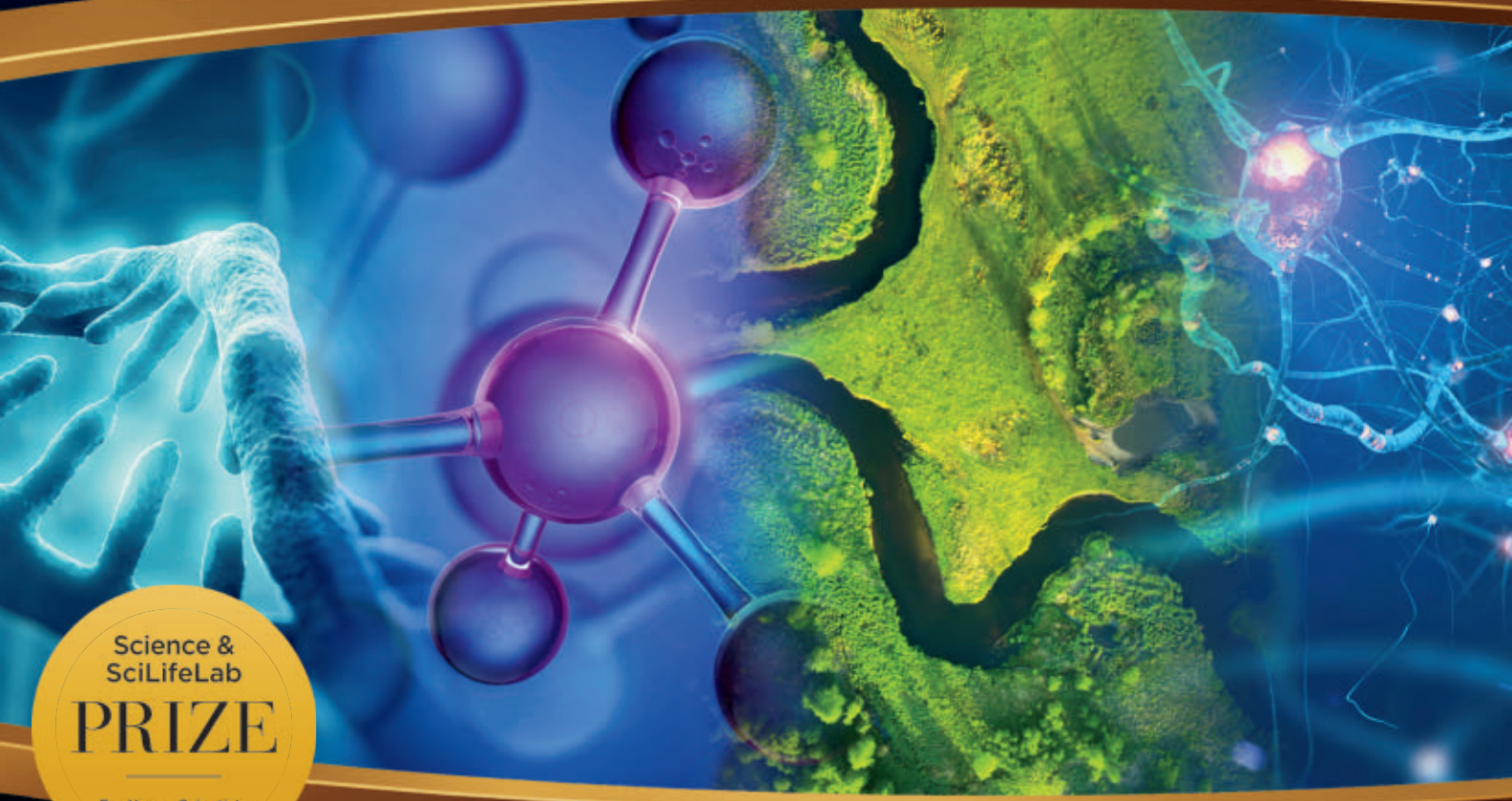
1. Center for Epitaxial van der Waals Quantum Solids, https://dmpl.postech.ac.kr.
2. J. Seo *et al.*, *Nature* **599**, 576–581 (*2021*).
3. Kim Group at POSTECH, Optoelectronic Devices and Materials Laboratory, https://jhklab.weebly.com.
4. H. Jeong *et al.*, *Scientific Reports* **9**, 5736 (*2019*).
5. J. Kim *et al.*, *J. Am. Chem. Soc.* **143**, 1399–1408 (*2021*).
6. G.-H. Lee Lab, POSTECH, http://ghleelab.postech.ac.kr.
7. G.-H. Lee *et al.*, *Nature* **586**, 42–46 (*2020*).
8. E. D. Walsh *et al.*, *Science* **372**, 409–412 (*2021*).

Sponsored by

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Congratulations on your Ph.D.!
Ready for your next step?

Would you like to see your research published in *Science*, have the opportunity to receive USD 30,000, and get to travel to Sweden in December?



The *Science* & SciLifeLab Prize for Young Scientists is an annual prize awarded to early-career scientists. The prize is awarded in four research categories:

- Cell and Molecular Biology
- Ecology and Environment
- Molecular Medicine
- Genomics, Proteomics, and Systems Biology

Entrants for the 2022 prize must have received their Ph.D. between January 1, 2020 and December 31, 2021.

Applicants will submit a 1,000-word essay that is judged by an independent editorial team organized by the journal *Science*. Essays are assessed on the quality of research and the applicants' ability to articulate how their work would contribute to their scientific field.
Apply before July 15, 2022.

Knut och Alice Wallenbergs Stiftelse

SciLifeLab

Science AAAS

scienceprize.scilifelab.se

CONTENTS

13 MAY 2022 • VOLUME 376 • ISSUE 6594

685
Some climate models predict future warming that most experts think is implausible.

NEWS

IN BRIEF

676 News at a glance

IN DEPTH

679 New subvariants are masters of immune evasion
Vaccines and prior infection still prevent severe disease from new SARS-CoV-2 strains
By G. Vogel

680 Better lipids to power next generation of mRNA vaccines
New delivery systems aim to increase vaccine potency and reduce side effects
By E. Dolgin
PODCAST

682 Wrestling with bird flu, Europe considers once-taboo vaccines
To lessen the toll of culling, some countries launch vaccine trials despite trade implications and public health risks
By E. Stokstad

683 Pandemic delays continue to plague polar science
Backlog of NSF-funded work will allow few new projects *By P. Voosen*

684 Odds for winning NSF grants improve as competition eases
Agency reports a 17% drop in number of grant applications since 2011, troubling some observers *By J. Mervis*

685 'Hot' climate models exaggerate Earth impacts
U.N. report authors say researchers should avoid suspect models *By P. Voosen*

FEATURES

686 The bias hunter
Itiel Dror is determined to reveal the role of bias in forensics, even if it sparks outrage
By D. Starr

INSIGHTS

PERSPECTIVES

692 The limits of forest carbon sequestration
Current models may be overestimating the sequestration potential of forests
By J. K. Green and T. F. Keenan
REPORT p. 758

694 Immune cells impede repair of old neurons
Interfering with age-related neuroimmune interactions promotes nerve regeneration
By N. Gaudenzio and R. S. Liblau
RESEARCH ARTICLE p. 715

695 Mapping cell types across human tissues
Single-cell analyses reveal tissue-agnostic features and tissue-specific cell states
By Z. Liu and Z. Zhang
RESEARCH ARTICLES pp. 711, 712, 713, & 10.1126/SCIENCE.ABO0510

697 Recording bacterial responses to changes in the gut environment
A CRISPR-based tool reveals intestinal microbiota gene expression through time
By L. Zahavi and E. Segal
RESEARCH ARTICLE p. 714

698 Beating natural proteins at filtering water
Artificial fluorous channels outperform aquaporins in water permeation *By Y. Shen*
REPORT p. 738

699 Anomalous fluid flow in quantum systems
New regimes of hydrodynamics are probed with synthetic quantum matter
By A. Morningstar and W. Bakr
RESEARCH ARTICLES pp. 716 & 720

POLICY FORUM

701 Investor-state disputes threaten the global green energy transition
Global action on climate change could generate upward of $340 billion in legal claims from oil and gas investors
By K. Tienhaara et al.

BOOKS *ET AL.*

704 Rethinking genetic disease
Carriers of the fragile X premutation can experience medical conditions of their own
By B. Sullivan

705 The becoming of the human brain
A neuroscientist traces the development of the body's most complex organ
By R. D. Fields

PHOTO: UGURHAN/ISTOCK

HUMAN VACCINES PROJECT
DECODING THE IMMUNE SYSTEM

MICHELSON MEDICAL RESEARCH FOUNDATION

Boost Your Career, Disrupt the Status Quo

APPLY NOW
MICHELSON PRIZES: NEXT GENERATION GRANTS

The Michelson Prizes: Next Generation Grants support young investigators applying disruptive research concepts to significantly advance the development of vaccines and immunotherapies for major global diseases.

A rigorous and competitive international search will identify the most innovative projects. We encourage applications from the full spectrum of scientific disciplines related to human immunology, vaccine, and immunotherapy research.

Are you 35 or younger?
Apply now for the 2022 Michelson Prizes!

humanvaccinesproject.org/michelson-prizes
#michelsonprizes

Dr. Camila Consiglio
Michelson Prize Winner 2021
Karolinska Institutet

"The support from the Michelson Prizes will help me to investigate how diseases develop differently in males and females – and ultimately develop vaccine strategies optimized for biological sex."

APPLY TODAY

APPLICATIONS DUE:
JUNE 26, 2022

GRANT AWARD:

$150,000

LETTERS

706 India must protect Himalayan headwaters
By A. Sharma et al.

706 Addressing patent gender disparities
By J. Goodman

707 Advancing racial equity in Brazil's academia
By W. Oliveira et al.

707 Techincal Comment abstracts

RESEARCH

IN BRIEF

708 From *Science* and other journals

RESEARCH ARTICLES

Human cell atlas

711 The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans *The Tabula Sapiens Consortium*
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABL4896

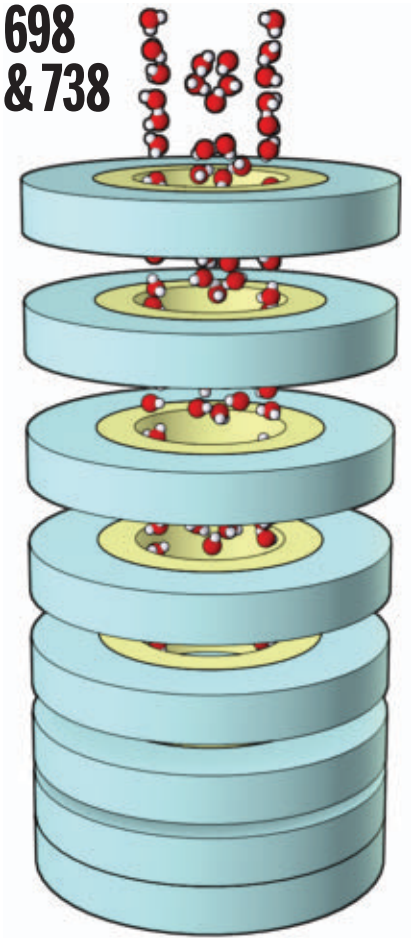
712 Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function *G. Eraslan* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABL4290

713 Cross-tissue immune cell analysis reveals tissue-specific features in humans *C. Domínguez Conde* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABL5197

PERSPECTIVE p. 695; RESEARCH ARTICLE 10.1126/SCIENCE.ABO0510

714 Microbiota
Noninvasive assessment of gut function using transcriptional recording sentinel cells *F. Schmidt* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABM6038
PERSPECTIVE p. 697

715 Neuroimmunology
Reversible CD8 T cell–neuron cross-talk causes aging-dependent neuronal regenerative decline *L. Zhou* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD5926
PERSPECTIVE p. 694



Quantum simulation

716 Quantum gas microscopy of Kardar-Parisi-Zhang superdiffusion *D. Wei* et al.

720 Observing emergent hydrodynamics in a long-range quantum magnet *M. K. Joshi* et al.
PERSPECTIVE p. 699

724 Neuroscience
Paradoxical somatodendritic decoupling supports cortical plasticity during REM sleep *M. Aime* et al.

731 Ferroelectrics
Highly enhanced ferroelectricity in HfO_2-based ferroelectric thin film by light ion bombardment *S. Kang* et al.

REPORTS

738 Nanochemistry
Ultrafast water permeation through nanochannels with a densely fluorous interior surface *Y. Itoh* et al.
PERSPECTIVE p. 698

744 Wetland ecology
High-resolution mapping of losses and gains of Earth's tidal wetlands *N. J. Murray* et al.

749 Organic chemistry
Modular access to substituted cyclohexanes with kinetic stereocontrol *Y. Li* et al.

754 Biomechanics
Recovery mechanisms in the dragonfly righting reflex *Z. J. Wang* et al.

758 Forest ecology
Cross-biome synthesis of source versus sink limits to tree growth *A. Cabon* et al.
PERSPECTIVE p. 692

762 Solar cells
Scalable processing for realizing 21.7%-efficient all-perovskite tandem solar modules *K. Xiao* et al.

DEPARTMENTS

675 Editorial
It ain't over 'til it's over
By H. H. Thorp

770 Working Life
Second chances *By Y. Xu*

ON THE COVER

A survey of cell types across tissues as part of the Human Cell Atlas, mapped with single-cell transcriptomics in three papers in this issue, lays the foundation for understanding how cellular composition and gene expression vary across the human body in health, and for understanding how genes act in disease. See pages 695 and 711–713.
Illustration: N. Fuller & C. Agosti of SayoStudio

Science Careers 768

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices.

Copyright © 2022 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS.

Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2212; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

CREDITS: (GRAPHIC) V. ALTOUNIAN/*SCIENCE*; (ILLUSTRATION) ROBERT NEUBECKER

2022 Mike Hogg Award and Lecture

Recognizing May-Britt Moser, Ph.D.

MD Anderson Cancer Center congratulates May-Britt Moser, Ph.D., on being named the 2022 Mike Hogg Award recipient.

With this honor, MD Anderson recognizes Moser's research, together with long-term collaborator Edvard Moser, on fundamental cognitive functions shared across animals, with a mechanistic focus on how the brain calculates self-position and how the information is stored in memory.

Moser is a professor of Neuroscience, founding director of the Center for Neural Computation, and co-director of the Kavli Institute for Systems Neuroscience at the Norwegian University of Science and Technology. She was awarded the Nobel Prize in Physiology or Medicine in 2014, together with E. Moser and John O'Keefe.

Moser's lecture will be delivered virtually. The event is sponsored by the Mike Hogg Fund and hosted by MD Anderson's Division of Education and Training.

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

Where Science Gets Social.

AAAS.ORG/COMMUNITY

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

EDITORIAL

It ain't over 'til it's over

H. Holden Thorp
Editor-in-Chief,
Science journals.
hthorp@aaas.org;
@hholdenthorp

The Biden administration is sheepishly waving a checkered flag on the pandemic. If you look closely, you can see its members cringing as they do so. Chief Medical Advisor Anthony Fauci told the PBS Newshour that the United States was "out of the pandemic phase" and then walked it back, saying he meant that the "acute component" of the pandemic phase was over. President Biden attended the likely superspreading White House Correspondents' Dinner last weekend but skipped cocktails and the meal, opting to just give his talk. Fauci avoided the whole affair. Meanwhile, Vice President Harris continued to isolate after her positive COVID-19 test, and many members of Congress and the administration announced positive test results as well. All of this happened while the White House allowed a renegade federal judge in Florida (where else?) to end the nationwide mask mandate without much of a fight. These mixed messages have been emanating from the administration for months now, and although those with resources have tools to manage COVID-19, care needs to be taken that those without such means are not forgotten.

> "... more variants... are on the horizon... this is no time to drop the ball."

When Biden pledged to "follow the science," it was hard to imagine that the country could have ended up here. But the administration made a big bet that vaccines would provide sterilizing immunity and end the pandemic, allowing it to move on to other priorities. Leaving behind the insanity of ivermectin, hydroxychloroquine, and bleach was certainly a great step forward. However, evolution has had other plans, and variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, the virus that causes COVID-19) have kept the pandemic going. This left the White House in a very tight spot: There was little political will to keep pushing nonpharmaceutical interventions, yet the pandemic was far from over. Add to this mounting inflation worries and concerns about the war in Ukraine, and the response has been a clumsy pivot to a message that politicians always turn to: personal responsibility. Get vaccinated, get boosted, wear a mask, get a prescription for the antiviral Paxlovid—if you want to. This may be fine if you have a healthy immune system, great health insurance, and the ability to navigate the US health care system. But what about everyone else?

COVID-19 is at a similar place to where the HIV/AIDS global pandemic was when the antiretroviral drugs came along. Yale epidemiologist Gregg Gonsalves told me about important parallels between both pandemics. "The HIV epidemic didn't go away," he said. "It just went to where people could ignore it. It went into the rural South, it went to communities that were already facing disparities in health." At that time, confusion between medicine and public health was also an important factor. "The discourse shifting to private choice and private adjudication of risk is really not what public health science is," he said. "We work in populations. And if we're talking about medicine, it's about private risk and private choices."

This all hit home for me when—while I was preparing this editorial—I tested positive for COVID-19. After writing about the virus for two-and-a-half years, it was in my body. But I'd had four shots of the vaccine to bolster my already robust immune system, a supply of rapid test kits, and a prescription for Paxlovid from my doctor. The fever was gone within a few hours of taking the antiviral, and I tested negative a few days later. Great news for me, but not for those who don't have these resources.

Legendary public health leader Paul Farmer summed up this situation well: "Those whose lives are rarely touched by structural violence are uniquely prone to recommend resignation as a response to it," he said. "In settings in which all of us are at risk, as is sometimes true of contagion shared through the air we breathe, we must also contemplate containment nihilism—the attitude that preventing contagion simply isn't worth it."

SARS-CoV-2 is rapidly mutating and recombining, and more variants and subvariants—potentially more pathogenic—are on the horizon. The world is still barely vaccinated, and even in wealthy countries like the United States, resources are inequitably distributed. It absolutely ain't over. And this is no time to drop the ball.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

Published online 5 May 2022; 10.1126/science.abq8460

NEWS

> “We have to kind of hold hope and grief at the same time.”
>
> **Jacquelyn Gill of the University of Maine, Orono,** quoted in the Associated Press on the mind frame of climate scientists.

IN BRIEF

Edited by **Lila Guterman**

SEISMOLOGY

A martian big one

NASA reported this week that its InSight lander on Mars hit the seismic jackpot, recording a magnitude 5 marsquake, by far the largest ever seen. Since InSight landed in 2018, seismologists have dreamed of such a quake, large enough to allow waves to encircle the planet’s surface, providing InSight’s seismic station an ultraprecise epicenter location and with it, a sort of skeleton key for the planet’s interior. Using smaller quakes, InSight’s team has managed to chart the thickness of the martian crust, mantle, and core; with the new quake, this picture is expected to grow far more precise. The quake was well timed, coming 3 days before the lander entered a hibernating “safe mode” because of low power from its dust covered solar panels. The spacecraft is likely to expire in the coming weeks as martian winter approaches.

U.S. limits use of J&J vaccine

COVID-19 | The COVID-19 vaccine made by Johnson & Johnson (J&J) should only be given to people who cannot receive any other vaccine, the U.S. Food and Drug Administration (FDA) said on 5 May. The move is related to thrombosis with thrombocytopenia syndrome (TTS), a rare, serious clotting disorder that has been associated with the vaccine. FDA has confirmed 60 cases of TTS, nine of them fatal, after 18.7 million doses of J&J were administered. The vaccine’s benefits still outweigh the risks, the agency says, but alternatives that don’t cause TTS are readily available. Scientists have not been able to pin down why the vaccine—and a similar one, made by AstraZeneca—cause the syndrome.

Verdict in China Initiative case

RESEARCH SECURITY | An applied math professor at Southern Illinois University (SIU), Carbondale, last week was found guilty of failing to report a Chinese bank account on his U.S. tax returns but cleared of charges that he committed grant fraud. The 4 May verdict against Minqqing Xiao came in the fourth jury trial resulting from the China Initiative, a controversial U.S. law enforcement campaign that has led to the prosecution of some two dozen U.S. academics, most of them of Chinese ancestry. “It’s a massive victory for Ming because the government was not able to prove that Ming did anything wrong in applying for his [National Science Foundation] grant,” said Ed Benyas, an SIU music professor who helped organize daily vigils to the Benton, Illinois, courthouse, where supporters wore “I stand with Ming” buttons. Xiao, on paid administrative leave from SIU, faces up to 5 years in prison and a substantial fine. Sentencing is set for 11 August.

Donor pig heart harbored virus

BIOTECHNOLOGY | The pig heart transplanted into a human patient in a groundbreaking surgical experiment carried a porcine virus that may have contributed to the man’s death, *MIT Technology Review* reported last week. In January at the University of Maryland School of Medicine (UMSOM), David Bennett, a heart failure patient, received the heart of a genetically modified pig created by the company Revivicor. When UMSOM researchers announced Bennett’s death in March, they shared no details about its cause. But in a 20 April webinar, UMSOM surgeon Bartley Griffith said the team had detected porcine cytomegalovirus in Bennett’s blood. The virus has been shown to damage pig hearts transplanted into baboons, leading to the primates’ death. The finding suggests immune rejection of the organ was not to blame for Bennett’s death, and that thorough virus screening before transplant could improve survival time.

Volunteers, AI bag 1000 asteroids

PLANETARY SCIENCE | An army of 11,400 citizen scientists joined forces with artificial intelligence (AI) to spot more than 1000 previously unknown asteroids. In 2019, researchers led by the European Space Agency enlisted volunteers to peruse more than 37,000 images taken by the Hubble Space Telescope over nearly 20 years. Because the exposures were 30 minutes long, asteroids appear as

Asteroid trails appear curved or S-shaped in this image from the Hubble Space Telescope.

PHOTO: NASA; ESA; B. SUNNQUIST AND J. MACK/STSCI

This bridal veil stinkhorn fungus breaks down plant matter in the soil.

PHOTO: ALEX HYDE/NPL/MINDEN PICTURES

ECOLOGY

Soil fungi keep ecosystems humming

Mushrooms not only add color to the landscape, they help keep the natural world productive and stable. Places with a wider variety of soil fungi that break down plant matter—the so-called decomposers—are better able to cope with drought and other stressors, ecologists reported this week in *Nature Ecology & Evolution*. The researchers determined the fungal makeup of about 700 soil samples from tropical, temperate, and polar climates. They mapped the diversity of different kinds of soil fungi onto 2 decades' worth of satellite images of plant photosynthesis, a measure of the ecosystem's productivity. Diverse fungal decomposers helped plants stay productive over time, but diversity of pathogenic fungi had the opposite effect, says lead author Manuel Delgado-Baquerizo, an ecosystem eco-logist at the Spanish National Research Council. He and others suggest finding a way to enhance decomposer diversity may buffer against the negative effects of climate change.

curved lines or streaks. The volunteers spotted more than 1000 of these traces, which were then used to train an AI system to find more, bumping the total to 1701 in 1316 images, the team announced last week. Roughly one-third are known asteroids, but the rest are new, and are thought to be small lumps of rock in the main belt between Mars and Jupiter. Researchers will now try to assess their orbits, sizes, and rotation rates, to learn more about the rubble from which the planets formed.

Disgraced surgeon on trial

MISCONDUCT | Paolo Macchiarini, the surgeon famed and then disgraced for implanting artificial tracheae seeded with stem cells into patients, took the stand in his trial for assault in Solna, Sweden, last week. Macchiarini performed at least eight of the transplants between 2011 and 2014. All recipients died except one, who had the implant removed. In 2016, the Karolinska Institute (KI) fired Macchiarini, who was later found guilty of fraud and research misconduct. Now, Swedish prosecutors have charged him with assault or "bodily harm due to negligence" related to three patients who received transplants at KI. Prosecutors argue Macchiarini performed the transplants with "reckless intent," knowing complications plagued earlier recipients. But Macchiarini testified that he performed the surgeries with the knowledge and support of KI and that the surgeries were a last resort for the patients. The trial, which began on 27 April, is expected to conclude on 23 May.

Stanford gets climate windfall

PHILANTHROPY | Stanford University will use a $1.1 billion gift—the second largest to an academic institution in history—to create a school focused on climate change. The gift, from John Doerr, a Silicon Valley venture capitalist, and his wife Ann Doerr, board chair of Khan Academy, will establish a school with 90 faculty members, including many from existing departments; Stanford will add 60 more over a decade. Climate-focused schools are a trend in U.S. higher education, with Columbia University announcing a similar one in 2020.

Pediatricians take stand on race

BIOMEDICINE | The American Academy of Pediatrics (AAP), a professional association, announced last week it would eliminate race-based medicine in its recommendations and publications. Race has no use as a proxy for a person's risk of health conditions or for other biological traits, AAP said in its policy announcement, noting that race-based medicine has increased racial health disparities. The group will revisit its "entire catalog," including its guidance for newborns with jaundice, which currently says that East Asian race is a risk factor for the condition.

U.S. preps for quantum hackers

CRYPTOGRAPHY | Someday, quantum computers—exotic machines that can solve practical problems that would stymie conventional computers—will be able to crack the encryption algorithms used to protect internet communications. So, last week President Joe Biden's administration instructed federal agencies to prepare to shift to new algorithms that can resist a quantum attack. The memo envisions

the decadelong shift beginning in 2024, when the first standard for such "post-quantum cryptography" should be set by the National Institute of Standards and Technology (NIST). Fortunately for internet companies, postquantum cryptography will involve changes mostly in software. Still, NIST researchers say the transition will be challenging and expensive.

NSF revs up innovation engines

TECHNOLOGY TRANSFER | Details of the competition for the largest cash awards in National Science Foundation (NSF) history were announced last week. NSF hopes to spend $160 million over 10 years on each of five "regional innovation engines" that will headline the agency's new technology directorate. The competition kicks off this year as teams of university scientists, companies, and local governments vie for 50 $1 million starter grants. NSF is asking applicants for their plans to conduct user-inspired research, translate the results into high-tech products and services, and train local people for the additional jobs that are created. The five most promising approaches will win the jackpot in a second round of funding next year. NSF says it will favor teams from areas "without well-established innovation ecosystems," meaning long odds for traditional high-tech regions like Silicon Valley, Boston, and North Carolina's Research Triangle Park.

$334M penalty in patent case

INTELLECTUAL PROPERTY | Last week, a Delaware jury ordered U.S. sequencing company Illumina, Inc. to pay $333.8 million to Chinese genomics firm BGI Group after finding Illumina infringed two of BGI's patents for DNA sequencing technology. Illumina and Complete Genomics, Inc., a California-based arm of BGI, have each claimed the other infringed their "2-channel" technology patents, which speed sequencing by imaging bases in parallel. Illumina has said it plans to appeal the verdict. Meanwhile, the companies are also locked in patent battles in Denmark, Germany, Switzerland, and Turkey. The decision comes on the heels of another case in San Francisco, where a jury found BGI infringed a different Illumina patent and ordered it to pay $8 million.

U.S. gun murders surged in 2020

PUBLIC HEALTH | Firearm homicide rates in the United States increased by 35% from 2019 to 2020, reaching their highest level since 1994, the U.S. Centers for Disease Control and Prevention reported this week in the *Morbidity and Mortality Weekly Report*. The rate of gun murders increased from 4.6 to 6.1 per 100,000 people as the COVID-19 pandemic took hold. The biggest increases were in Black boys and men ages 10 to 44 and in American Indian/Alaska Native men ages 25 to 44. Rates also grew with increasing poverty. The firearm suicide rate also grew in 2020, by 2.5%; among American Indian/Alaska Natives, the rate grew by 41.8% even as gun murder rates in this group grew by 26.5%, less than the overall U.S. increase.

Starlings fly in enormous, intricate flocks called murmurations.

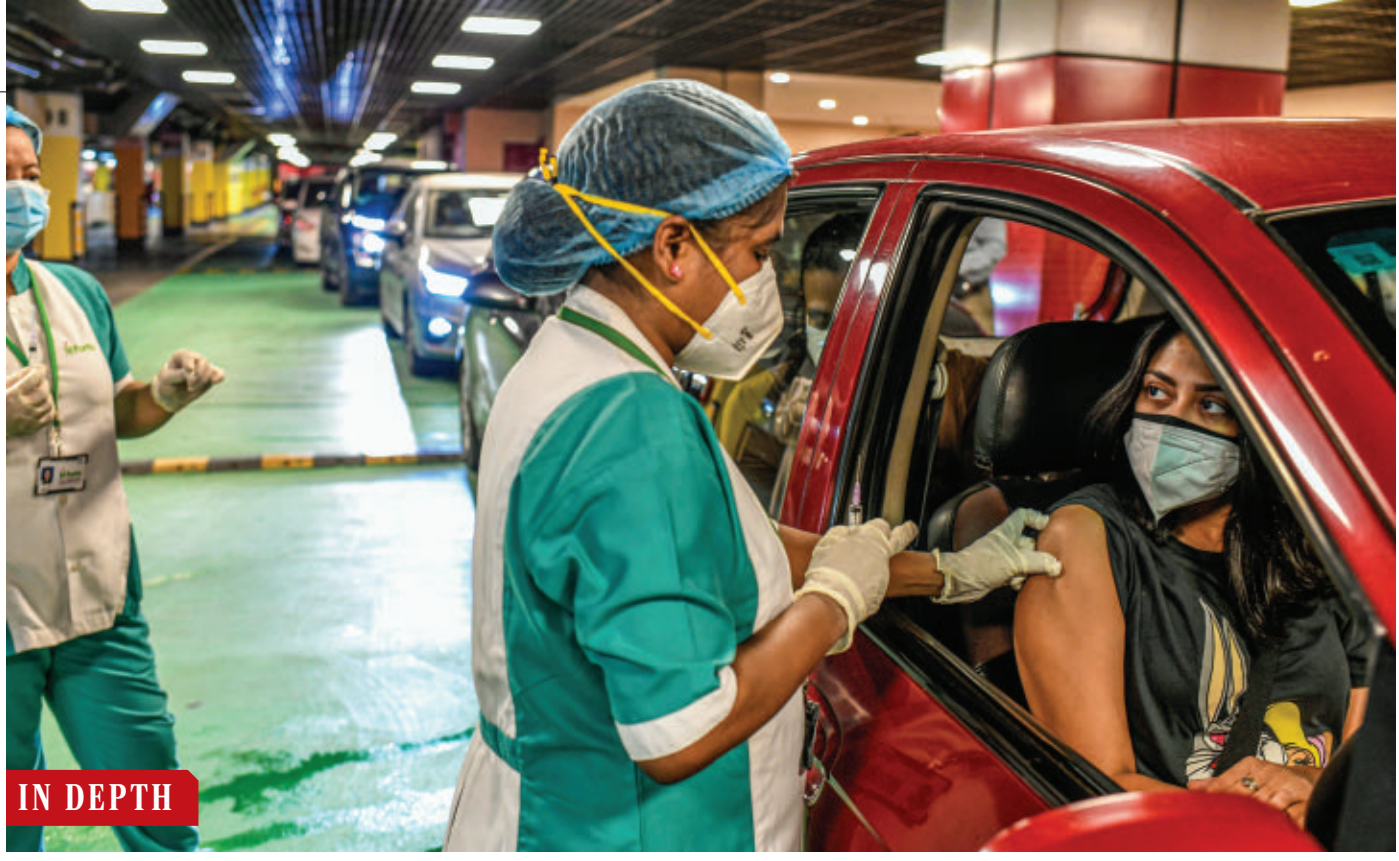
PHOTO: LAURIE CAMPBELL/MINDEN PICTURES

BIOLOGY

Speed control keeps flocks in sync

Large flocks of starlings doing their synchronized dances across the sky stay together without constantly bumping by adjusting to one another's subtle changes in speed, researchers have determined. Theoretical physicist Antonio Culla from the Italian National Research Council and colleagues recorded flocks of starlings and then experimented with computer models until they came up with one whose on-screen birds behaved like real birds. The secret, Culla's team reported this week in *Nature Communications*, is having "marginal speed control": The birds, which fly 8 to 18 meters per second, slow down or speed up a little to keep up with birds nearby but avoid large changes in speed, which could lead to the breakup of the flock, he says. The finding could one day help drones better coordinate their flight in swarms.

SCIENCE.ORG/NEWS
Read more news from *Science* online.



IN DEPTH

A woman receives a COVID-19 vaccine in New Delhi in June 2021. The continuing evolution of SARS-CoV-2 is a challenge for vaccine developers.

PHOTO: ATUL LOKE/PANOS PICTURES/REDUX

COVID-19

New subvariants are masters of immune evasion

Vaccines and prior infection still prevent severe disease from new SARS-CoV-2 strains

By **Gretchen Vogel**

Once again, South Africa is at the forefront of the changing COVID-19 pandemic. Epidemiologists and virologists are watching closely as cases there rise sharply again, just 5 months after the Omicron variant caused a dramatic surge. This time, the drivers are two new subvariants of Omicron named BA.4 and BA.5, which the Network for Genomic Surveillance in South Africa first detected in January.

The new strains didn't have much of an impact initially, but over the past few weeks case numbers in South Africa jumped from roughly 1000 per day on 17 April to nearly 10,000 on 7 May. A third subvariant called BA.2.12.1 is spreading in the United States, driving increases along the East Coast.

It's still unclear whether the new subvariants will cause another global COVID-19 wave. But like the earlier versions of Omicron, they have a remarkable ability to evade immunity from vaccines, previous infection, or both—a disturbing portent for the future of the pandemic and a potentially serious complication for vaccine developers.

In most cases, vaccination or earlier infection still seem to provide protection from severe disease. "There's no reason to freak out," says John Moore, an immunologist at Weill Cornell Medicine. The new strains are "an additional hassle," he says, but "there's no indication that they're more dangerous or more pathogenic."

Hospitalizations in South Africa, for example, have increased, "but because it is starting from a very low level, it's not cause for alarm," says virologist Tulio de Oliveira of Stellenbosch University, who helped identify BA.4 and BA.5. Numbers of patients in intensive care units are as low as they have been since the start of the pandemic, he says. "At the moment, we expect something similar to the Omicron BA.1 wave," when hospitalization rates stayed manageable.

The new superspreaders do, however, showcase the restless virus' ability to find ways around the "immunity wall" built up over the past 2 years and to continue to circulate at high levels. Even if the new variants cause relatively little severe disease, "it's a numbers game," says Leif Erik Sander, an infectious disease expert at the Charité University Hospital in Berlin; enough new infections could still overwhelm health systems.

All three new strains share key mutations with the BA.2 strain of Omicron, which, like BA.1, emerged in southern Africa in October 2021. Initial studies by de Oliveira and Alex Sigal, an infectious disease expert at the Africa Health Research Institute in Durban, suggest BA.4 and BA.5 can elude the immunity of patients who were infected with the BA.1 strain, which in South Africa caused a much larger wave than BA.2. That may be in part because immunity has waned since South Africa's BA.1 wave peaked in December. People who were both vaccinated and infected had somewhat stronger protection, de Oliveira and Sigal reported in a 2 May preprint.

All three new variants have mutations that alter a key amino acid called L452, which may help explain their ability to dodge immunity. L452 is part of the receptor-binding domain, the part of the spike protein that locks onto cells, enabling infection. The domain is also a key target for protective antibodies.

The Delta variant that caused devastating surges around the world in 2021 had mutations in L452 as well, so many scientists

have been watching this hot spot carefully, including immunologist Yunlong Richard Cao of Peking University. On 11 April, Cao says, he and his colleagues noticed a pattern: New Omicron sublineages from New York, Belgium, France, and South Africa all had changes in L452. "The independent appearance of four different mutations at the same site? That's not normal," Cao says. The researchers suspected it was the virus' response to the high levels of immunity generated by the huge Omicron waves.

They immediately started to make copies of the spike protein based on the new sequences and test how well different antibodies could block those proteins, preventing them from binding to cells. They used sera from 156 vaccinated and boosted subjects, including some who had recovered from either BA.1 or severe acute respiratory syndrome (SARS), the coronavirus disease that caused a deadly global outbreak almost 2 decades ago. Like the South African team, they found that blood from patients who had been infected with BA.1 had only weak ability to neutralize BA.4 and BA.5; the same was true for BA.2.12.1. Even less effective were sera from people who had previously been infected with SARS and then vaccinated against COVID-19, they reported in a 2 May preprint.

The latter finding was surprising. Previous work by Linfa Wang, a bat coronavirus researcher at the Duke-NUS Medical School in Singapore, had shown patients who had recovered from SARS and were then vaccinated had strong protection against earlier SARS-CoV-2 variants—and even some related animal viruses—a finding that seemed to hold clues to developing vaccines effective against multiple coronaviruses, including those that might trigger the next pandemic. But the new mutations apparently helped the Omicron subvariants evade those previously powerful antibodies.

Wang notes, however, that the subjects in the new study were all vaccinated with CoronaVac, a Chinese vaccine made from inactivated virus. Subjects in his study were vaccinated with messenger RNA (mRNA) vaccines, which might provide a more potent response to the new strains, he says. But Wang agrees that Omicron's knack for immune escape is dramatic. Based on its immunological profile, it "should be called SARS-3," he says—an entirely distinct virus.

Omicron's rapid evolution creates difficult decisions for vaccine- and policymakers about whether to shift to a new set of vaccines or stick with the current formulations, which are based on the virus that emerged in Wuhan, China, more than 2 years ago. Moderna has tested two "bivalent" versions of its mRNA vaccine, containing the ancestral strain and either the Beta variant—which spread in South Africa for a while in 2021 but is now gone—or the Omicron BA.1 variant. The company has not yet reported data on how well they might protect against the new subvariants.

Pfizer, the other mRNA vaccine producer, has tested the efficacy of a booster and a primary vaccine based on BA.1. Results are expected by the end of June. The U.S. Food and Drug Administration has scheduled a meeting for 28 June to analyze available data and make vaccine recommendations for the fall.

The limited protection that BA.1 infection provided against the new subvariants in lab studies has already raised questions about how useful the new Omicron-specific vaccines might be. Wang says the virus is evolving too quickly for strain-specific vaccines to keep up. Instead, a broad cocktail of monoclonal antibodies targeting different strains might be the best way forward, he says.

> **"We need to focus on broadening our immunity."**
> **Kristian Andersen**, Scripps Research

Such a shot could prevent infections for several months in those vulnerable to severe disease, including immunocompromised people who don't respond to vaccines. Protecting that group is crucial, he notes, because many researchers suspect new variants emerge during long-term infections in people whose immune systems fail to clear the virus. The main hurdle, Wang says, is cost: A dose of monoclonal antibodies is about $1000 per patient, he notes, "but if someone could find a way to lower that to $50 or $100," the approach could be cheaper than constantly updating vaccines.

Kristian Andersen, who studies viral evolution at Scripps Research, draws a sobering lesson from the newest Omicron variants. Although we don't know what future variants will look like, he says, "we can be certain that they'll continue to be more and more capable of immune escape," possibly leading to lower protection against not just infection, but also against severe disease. "We need to focus on broadening our immunity," he says.

It's far from clear what kind of vaccine might prompt that broadened immunity, but "we really, really need to get going" to figure that out, Andersen says. "Simply letting the virus do what viruses do—continue to infect us, and likely several times a year—just isn't an option in my playbook." ■

VACCINE DEVELOPMENT

Better lipids to power next generation of mRNA vaccines

New delivery systems aim to increase vaccine potency and reduce side effects

By **Elie Dolgin**

As any dietician will tell you, some fats are good—and that is surely true of the little fatty balls found in two of the world's most widely used COVID-19 vaccines. Known as lipid nanoparticles (LNPs), these tiny bubbles of fat encase messenger RNA (mRNA) that encodes a viral protein, helping ferry it into cells and shield it from destructive enzymes. The technology was key to the success of COVID-19 shots from Moderna and the Pfizer-BioNTech collaboration. But as beneficial as these fats are, there is plenty of room for improvement.

The nanoparticles are a major source of unwanted side effects when they spread through the body, triggering the aches and inflammation many people experience after vaccination. They do a poor job of unloading their cargo once inside cells, a necessary step for the proteinmaking machinery to turn the mRNA sequences into immune-priming signals. And because they tend to fall apart when warm, they have to be stored at low temperatures, limiting their global use.

"This is a system that clearly has legs," says biochemist Pieter Cullis of the University of British Columbia (UBC), Vancouver, who created the first LNPs, but "we still need to increase the efficiency of LNPs—that's for sure."

A new generation of LNPs with greater potency, fewer side effects, increased stability, and more precise tissue-targeting properties is now under development at big pharma and biotech startups. Big money is at stake: These improved nanoparticles could lead to better mRNA vaccines for COVID-19 and other diseases. They might also help mRNA deliver on its promise as a therapeutic tool to treat disease. "There are innovations on delivery that certainly could

change the game," says Philip Santangelo, a biomedical engineer at the Georgia Institute of Technology who has collaborated with several mRNA companies.

Cullis and colleagues developed the first LNPs about 20 years ago to carry gene-silencing drugs into cells. He and others later tailored the LNPs' four lipid components to deliver disease-correcting mRNA to faulty cells. Now that they are being put to a new use, in vaccines, "There is still so much optimization and development that needs to happen," says UBC bioengineer Anna Blakney, co-founder of the RNA vaccine company VaxEquity. And when it comes to understanding how cells interact with the nanoparticles, "it's just this big question mark," she adds.

One clue emerged earlier this year when Genentech scientists showed how nanoparticles activate a particular inflammatory pathway, the interleukin-1 axis, which is critical to generating protective immune responses but can also spur side effects. Among the LNPs tested, one made with SM-102, an "ionizable" lipid that helps bind and package mRNA into LNPs, proved an especially strong instigator of this pathway. That could help explain why Moderna's shot, which relies on SM-102, is both highly effective and prone to making people feel icky.

The Genentech team did not evaluate the comparable lipid found in the Pfizer-BioNtech vaccine. But Mohamad-Gabriel Alameh and colleagues from the University of Pennsylvania Perelman School of Medicine tested a closely related one and found that it triggered a wide range of inflammatory molecules, both desired and not. The goal now is to design ionizable lipids that activate favorable immune pathways without overstimulating detrimental ones, says Alameh, who co-founded AexeRNA Therapeutics with bioengineer Michael Buschmann of George Mason University and others. "Is it very simple? No," Alameh says, but it should be possible.

Before his death in March, Buschmann led a team that in 2021 showed how the electric charge of LNPs is critical to vaccine success. A negative charge makes the particle less likely to stay in the muscle and lymph nodes of injected mice, where it could elicit beneficial immune responses; instead it tends to spread widely, raising the risk of fevers, chills, and other adverse reactions.

To make a less negatively charged nanoparticle, the researchers tweaked the chemistry of the ionizable lipid. When formulated into an mRNA vaccine for COVID-19, the new LNP carrier prompted mice to produce more protective antibodies than standard delivery systems and had fewer side effects, according to data posted last year in preprint form and now under review at *Nature Communications*.

Dan Peer, a biochemist at Tel Aviv University and co-founder of the vaccine delivery startup NeoVac, has also developed libraries of new ionizable lipids with atypical structures. In unpublished experiments, they seem to enable better mRNA vaccines with fewer side effects and also extend their shelf stability at room temperatures.

Other improvements could come from boosting the uptake of LNPs into cells and then enhancing their ability to break free of the sacs of cell membrane, known as endosomes, that carry them inside. The vast majority of LNPs get trapped in these receptacles and then destroyed or ejected without delivering their vaccine payloads, meaning "there's a huge amount of RNA that's not being used," says Gaurav Sahay, a bioengineer at Oregon Health & Science University.

The shape of ionizable lipids affects an LNP's ability to disrupt an endosome, as does cholesterol, one of the other fats in LNPs. Together with Moderna scientists, Sahay and colleagues reported in 2020 that using different forms of cholesterol can enhance rates of LNP escape from endosomal entrapment. He has founded a company called Enterx Biosciences to commercialize his discoveries.

Sanofi has begun to evaluate some of its customized LNPs head-to-head in human trials. In a study launched in 2021, for example, the company assessed two LNP options for delivering an mRNA flu shot it is developing. According to preliminary data, one lipid formulation proved much better at kick-starting anti-influenza immunity, Frank DeRosa, head of research and biomarkers at Sanofi's mRNA Center of Excellence, announced at an investor event in December 2021. But the same LNP also provoked more frequent side effects at higher doses.

Other firms, including BioNTech and Arcturus Therapeutics, have begun to explore ways to eliminate polyethylene glycol, a compound that helps stabilize LNPs but has also been linked to some types of bad vaccine reactions. Many more companies, meanwhile, are focused on optimizing lipids for delivering mRNA to treat disease rather than prevent it. This requires getting mRNAs that encode disease-correcting proteins to the precise cells and tissues where they are needed—not just to the liver, where current LNP formulations tend to end up following infusion. "Delivery of LNPs will be key to really expanding the reach of mRNA" beyond preventative vaccines, says Dominik Witzigmann, co-founder and chief executive of the Cullis-founded startup NanoVation Therapeutics.

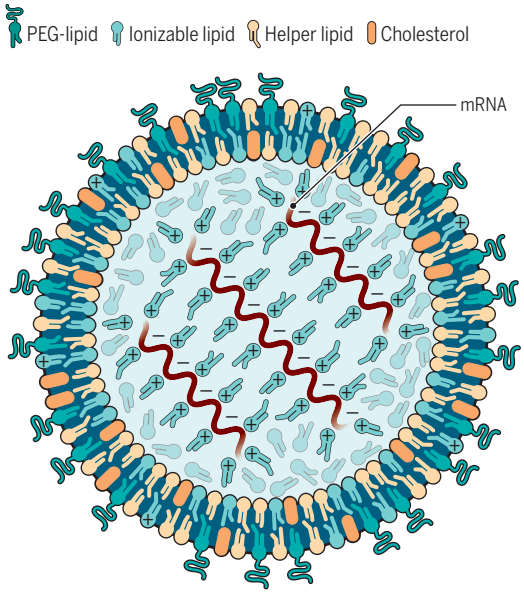
The heightened focus on LNP technologies, along with the profits reaped from the COVID-19 vaccines, has brought increased litigation. Alnylam, which helped develop the first approved medicine delivered in an LNP—a gene-silencing drug marketed since 2018 to treat a rare neurodegenerative disorder—claims that its foundational patents cover lipid components of the Moderna and Pfizer-BioNTech vaccines. And Arbutus BioPharma, yet another Canadian firm co-founded by Cullis, is seeking damages from Moderna for allegedly infringing on a patent that covers LNPs comprising certain ratios of lipids.

But these intellectual property disputes are unlikely to have a chilling effect on LNP innovation, says Jacob Sherkow, a biotech patent attorney with the University of Illinois College of Law. "There's too much money at play." ■

Elie Dolgin is a science journalist based in Somerville, Massachusetts.

The particulars of nanoparticles

To boost vaccine potency and limit side effects, researchers are altering each of the four ingredients that make up lipid nanoparticles. Each particle includes ionizable lipids that bind mRNA and shift their charge from positive to neutral once in the body to limit the particle's toxicity. The three other types of fats contribute to its structure and stability. Helper lipids also aid particles in fusing with cells, cholesterol helps them escape from cells' endosomes, and polyethylene glycol (PEG)-lipids prevent them from clumping to help prolong their action.



GRAPHIC: M.D. BUSCHMANN, M.J. CARRASCO, *ET AL.*, *VACCINES*. 2021; 9(1):65. ADAPTED BY A. MASTIN/*SCIENCE*

These turkeys in Denmark and millions of other birds in Europe were killed to stop an avian influenza virus.

PHOTO: MADS CLAUS RASMUSSEN/RITZAU SCANPIX/AFP/GETTY IMAGES

INFECTIOUS DISEASE

Wrestling with bird flu, Europe considers once-taboo vaccines

To lessen the toll of culling, some countries launch vaccine trials despite trade implications and public health risks

By **Erik Stokstad**

In March, Christian Drouin, a French farmer in the Vendée, discovered that his chickens were dying of avian influenza. He had to take drastic action to cull his flock and prevent the infection from spreading. Normally, veterinarians would arrive to gas the birds with carbon dioxide. But the veterinary teams were overwhelmed with calls to cull flocks infected with the virus, now apparently endemic in Europe.

So Drouin was advised to switch off the ventilation fans in his poultry buildings. As the temperature rose, most of his 18,000 birds died of heat stroke over several hours. The next day, his neighbors helped him bury the carcasses. "After that, I lay down in the dark, stunned by what I had done," he told Agence France-Presse.

Seeking to stamp out the highly pathogenic H5N1 strain of avian influenza, France and other countries have been culling record numbers of poultry—more than 16 million birds since December 2021 in France alone. Last year, the cost there exceeded €150 million. Now, faced with the desperation of farmers like Drouin, France, the Netherlands, and other hard-hit countries have restarted research into a solution long considered taboo: vaccinating flocks.

Ministers in France and other EU countries are discussing the idea, and Dutch scientists have already begun trials of chicken vaccines. In southwestern France this week, researchers will begin to immunize ducks with a newly developed vaccine. And in October, stakeholders will gather at the World Organisation for Animal Health to discuss how to lower international barriers to shipping vaccinated poultry.

For now, many countries refuse such shipments because they aren't confident countries with vaccinated birds have controlled the virus. Flu experts worry vaccination may not completely halt outbreaks, perhaps raising the long-term risk that the avian virus leaps to humans. And developing and administering vaccines will be costly.

For all these reasons, vaccination is the last resort, says avian pathologist Jean-Luc Guérin of the National Veterinary School of Toulouse. "We use this tool only if we admit that we cannot control the infection by classical ways." The United States has not authorized the use of avian influenza vaccines because of the trade implications, hoping culling will stop its current outbreak. The U.S. poultry industry is taking a wait-and-see approach. But in Europe the devastation wrought by the virus, and the cost and logistics of culling millions of birds, may be changing the calculus.

In places where the extremely infectious new strain of avian influenza has taken hold, vaccination "really has the capacity to make a huge difference," says Richard Webby, a virologist at St. Jude Children's Research Hospital who studies influenza in birds and other animals. In the long term, researchers say, living with H5N1 may require not just vaccines, but a restructuring of dense European poultry operations.

For 3 decades, ever more strains of avian influenza have been emerging in Asia. The current H5N1 strain arrived in Europe in 2021. It was first detected in the United States in January and continues to spread (*Science*, 29 April, pp. 441 and 459).

Some researchers are concerned that vaccinating, if not done carefully, will allow H5N1 to persist and continue to mix with strains in wild birds, with the risk that it might evolve to spread among people. The risk for the European Union and United States, although low, is probably the highest since H5N1 emerged 25 years ago, Webby says. "We really don't want this virus lurking around in poultry farms," adds Adel Talaat, an infectious diseases expert at the University of Wisconsin, Madison.

In an encouraging sign, vaccines have lessened the impact of recent outbreaks, at least in China. In 2017 the country began mandatory vaccination of poultry against an H7N9 strain that was able to spread to people. Vaccination slashed the prevalence of the virus in poultry and the number of human infections dropped to zero. That accomplishment "could be replicated everywhere," says virologist Hualan Chen of the Harbin Veterinary Research Institute, who developed the vaccines.

The campaign also paid off for Chinese farmers, who resumed producing broiler chickens, according to a cost-benefit analysis published in March in *Preventive Veterinary Medicine*. And the United States continued to accept Chinese poultry products, showing trade barriers are not insurmountable. Vaccines also benefit animal welfare by reducing the need to cull flocks, adds Chen, who "strongly" recommends vaccinating poultry against H5 strains.

Still, the virus strain now in Europe, H5N1, can be difficult to control with vaccines because it infects many species, including ducks, whereas H7N9 is chiefly a problem in

chickens, Webby says. And economics may favor culling for sporadic outbreaks.

But in hard-hit France, vaccine trials are starting. Two vaccines will be tested on ducks raised for foie gras. Ducks carrying bird flu are the "ultimate reservoir," Guérin says, because they can spread the virus for up to 15 days before showing symptoms. In the trials, birds will be vaccinated on farms, then exposed to the virus in a lab. The goal is to reduce the amount of virus circulating and so protect other poultry species.

One vaccine, Volvac Best, is made by Boehringer Ingelheim and already used in countries outside Europe, including Mexico and Egypt, to immunize chickens against Newcastle disease and H5N1. Ceva created the other vaccine, the first RNA vaccine to be tested in poultry, specifically for ducks. Results should be available by the end of the year, says Gilles Salvat, a veterinary health expert at the French Agency for Food, Environmental and Occupational Health & Safety. If the vaccines prove effective at lowering viral levels, Salvat hopes they might be ready for market by the end of 2023.

After farmers vaccinate flocks, they'll need to make sure the virus isn't circulating silently in any birds that were missed or didn't respond fully to a vaccine. They will need to swab birds and test for the virus, which can spread on boots, clothing, tires, and even the wind. Such measures will also reduce the risk of spread to humans or wild species, says Carol Cardona, a veterinarian and avian influenza specialist at the University of Minnesota, Twin Cities.

"Vaccines can help, but are not the golden bullet," says Ron Fouchier, a virologist at Erasmus University Medical Center. The approach may still require some culling, he says, because viruses will continue to evolve and may occasionally escape vaccines.

The current outbreak could be a game changer because the virus is spreading in many wild bird species. If it becomes endemic in the United States, too, then vaccination could become necessary, researchers say. The U.S. Department of Agriculture is modifying existing vaccines and testing new ones against the current H5N1 strain.

In the bigger picture, the Chinese poultry sector needs to prevent viruses from spilling to wild birds, Fouchier says. And European countries need to restructure to avoid having many farms with dense flocks close together, researchers say—an even bigger challenge than implementing vaccination.

Cardona says it could take years to optimize and approve vaccines, as well as devise a vaccination strategy and reassure trade partners. "What are we waiting for?" she asks. "We need to get working." ■

COVID-19

Pandemic delays continue to plague polar science

Backlog of NSF-funded work will allow few new projects

By **Paul Voosen**

In March 2020, staff at McMurdo Station, the main U.S. research base in Antarctica, thought the future was bright. Long-planned renovations had begun, including the replacement of decrepit dorms by glossy new lodges capable of housing more than 200 people. But then the pandemic struck, shutting down most of two summer field seasons at McMurdo and other polar research sites in Antarctica and Greenland. In some places the effects of that shutdown will linger for the rest of the decade, the National Science Foundation (NSF) announced last week, delaying projects and limiting access to one of the rarest resources in geoscience: time on the ice.

In Greenland, government entry restrictions kept most researchers away in the summer of 2021. Although NSF kept its high-altitude Summit Station running year-round, only minimal maintenance occurred, says Jennifer Mercer, NSF's Arctic section head. With nearly 1 meter of snow falling each year on the camp, it will require a lot of literal digging out. "We have a constant battle maintaining buildings above grade," Mercer says.

> **"We're saturated for a while in key logistics areas."**
> **Stephanie Short**, National Science Foundation

After the dig-out this summer, research in Greenland will be about back to normal. Not so in Antarctica, where "we're saturated for a while in key logistics areas," says Stephanie Short, NSF's head of Antarctic logistics. No work has been done on the McMurdo renovation for the past 2 years, and space in the old dorms had to be reserved to house possible COVID-19 cases, leaving the agency down by more than 200 beds. "To get back to full strength," Short says, "we need that lodging building."

For now research in Antarctica will prioritize ongoing projects that feature either heavy international participation—such as the International Thwaites Glacier Collaboration—or critical annual measurements, says Michael Jackson, head of Antarctic earth sciences at NSF. New starts will be biased toward projects led by early-career researchers. But some new projects will have to be delayed again, he says. "That's heartbreaking for us," he says.

One deferred project is a plan to drill into Hercules Dome, an expanse of ice 400 kilometers from the South Pole. Ice cores from the dome might record the last time the West Antarctic Ice Sheet collapsed in a slightly warmer climate more than 100,000 years ago, and help predict when it might happen next. When NSF agreed to fund the project in 2020, researchers thought drilling might begin by 2023. Now, 2025 is more likely, says Eric Steig, the project's principal investigator at the University of Washington, Seattle.

Steig says the pandemic hit an enterprise that was already stretched. "NSF is always planning on more projects [in Antarctica] than they are likely to be able to support, so even without COVID we always run into major delays." And despite the agency's plan to give priority to early-career projects, he says, many young researchers may be left in the cold, as projects led by senior researchers also support many early-career researchers. But there are no simple solutions, and "I have a lot of confidence in the NSF program managers," he says.

Despite the isolation of the pandemic, one NSF project built new bridges between U.S. and Greenlandic researchers. It seeks to help Greenlandic communities understand the effect of climate change on local sea levels, which counterintuitively are set to fall by as much as several meters by century's end, as land sheds ice weight and rebounds and the gravitational tug of the massive ice sheet on the surrounding ocean ebbs. During the pandemic, Greenlandic researchers kept working without the U.S. team, says Kirsty Tinto, a Columbia University geophysicist on the project. They sailed on seafloor mapping cruises and interviewed community leaders—hunters, fishers, city planners—about how they use the waterfront. The pandemic showed the joint project's resilience, Tinto says. "I don't like pandemics. I don't like global despair." But, she says, "I do like having my expectations confounded." ■

The National Science Foundation could speed its granting process with more RAPID awards, often given for studies of natural disasters such as this Alaska landslide.

PHOTO: MATT BORON/ALASKA DEPARTMENT OF TRANSPORTATION AND PUBLIC FACILITIES/ASSOCIATED PRESS

FUNDING

Odds for winning NSF grants improve as competition eases

Agency reports a 17% drop in number of grant applications since 2011, troubling some observers

By **Jeffrey Mervis**

Federal research agencies strive to fund a healthy percentage of the grant applications they receive, ensuring that scientists can pursue their best ideas. A steadily rising budget is their preferred method for maintaining a robust success rate. But a fall in applications can have the same effect.

A new report from the National Science Foundation (NSF) on its merit review system documents how falling demand has boosted success rates at the $8.5 billion research agency. Released late last month, the analysis shows the annual number of applications submitted to NSF has dropped by 17% over the past decade, falling from 51,562 in 2011 to 42,723 in 2020. Success rates jumped from 22% to 28% during the same period, even though the number of grants awarded increased by 8%. The rate rose even faster—from 19% to 28%—for the agency's standard research grants.

Those trends are most visible in NSF's biology directorate, where demand has tumbled by 50% over the decade and the chances of winning a grant have doubled, from 18% in 2011 to 36% in 2020. Officials credit the shifts to their decisions to eliminate fixed deadlines for submitting proposals and make other tweaks to the grantsmaking process designed to ease the workload on beleaguered program officers. The geology directorate, which piloted some of the same changes but never adopted them across the board, has seen a 28% decline in the number of proposals, with success rates rising from 31% to 42%. The pattern is similar for the engineering directorate, which dropped deadlines in 2018.

Higher success rates please researchers and administrators. "The ability of faculty to get funds for their research is the ultimate metric, and success rates tell you that," says Sarah Nusser, a statistician and former vice president for research at Iowa State University. She says she "never paid much attention" to how many proposals faculty were submitting. But some policymakers wonder whether fewer cutting-edge research proposals could affect the nation's ability to innovate.

"We're not pleased that there has been the decline," says Stephen Willard, a biotech executive who is a member of NSF's presidentially appointed oversight body, the National Science Board. "We're trying to identify the reasons, with the goal of turning it around."

James Olds, a neuroscientist at George Mason University who led the biology directorate from 2014 to 2018, the period when the deadline rules were revised, thinks the changes helped improve the quality of proposals. "Life scientists are always collecting more data to test their hypotheses," he says, "so I prefer to see researchers submit a proposal when it's ready scientifically, not because of the craziness of a 5 p.m. deadline."

A senior NSF official, however, isn't sure the changes in the application process explain the overall drop in proposals. "The merit review report helps identify interesting trends and this is one of them," says Erika Rissi of NSF's Office of Integrative Activities, which does the analysis each year. "But the report doesn't diagnose anything, so we don't know why."

Higher education officials say many factors influence whether a researcher submits a grant, including their readiness to participate in a new solicitation or their perception of the odds of getting funded. The COVID-19 pandemic, which caused massive disruptions on campus, could also play a role. (The report only covers operations through September 2020, which spans the first 6 months of the pandemic.)

The trends are not uniform across the agency. The computing directorate now gets 20% more proposals than a decade ago, for example, and a 40% increase in its budget has helped it maintain a success rate of 24%. Demand in physical sciences and math has been flat, and success rates have crept up from 27% to 30%. Both directorates used fixed deadlines for submitting proposals during the period.

NSF's analysis also reveals large differences in the trends by gender and race/ethnicity. The number of proposals from women has declined by 12% over the decade, for example, compared with a 21% drop for men. The number of proposals from Black and Asian investigators has shrunk by 27% and 28%, respectively, a much steeper decline than for the overall pool.

The data-rich report includes a suggestion from the science board for improving another crucial metric: the "dwell time" between the receipt of a proposal and notifying the applicant of a funding decision. The agency's goal is to complete action on 75% of applications within 6 months, but the report notes it has fallen short of that target for the past 4 years.

NSF might be able to do better, the board said, by funding more proposals that don't require outside vetting. It cited the RAPID awards, an NSF-wide program that makes quick-turnaround grants of up to $200,000 to study the impact of a natural disaster or a major societal upheaval like the current pandemic. NSF quadrupled the number of RAPID grants in 2020 after receiving $75 million for them in the first massive federal COVID-19 relief package. Even so, they only made up 2.7% of all NSF research proposals that year.

"We're not saying go easy on them," Willard says. "But this could help with dwell times." ■

CLIMATE SCIENCE

'Hot' climate models exaggerate Earth impacts

U.N. report authors say researchers should avoid suspect models

By **Paul Voosen**

One study suggests Arctic rainfall will become dominant in the 2060s, decades earlier than expected. Another claims air pollution from forest fires in the western United States could triple by 2100. A third says a mass ocean extinction could arrive in just a few centuries.

All three studies, published in the past year, rely on projections produced by some of the world's next-generation climate models. But many of the models have a glaring problem: They predict a future that gets too hot too fast (*Science*, 30 July 2021, p. 474). Although modelmakers acknowledge the problem, researchers who use the model projections to gauge the impacts of climate change are sometimes unaware. That has resulted in a parade of "faster than expected" results that threatens to undermine the credibility of climate science, some researchers fear.

Scientists need to get much choosier in how they use model results, a group of climate scientists argues in a commentary published last week in *Nature*. Researchers should no longer simply use the average of all the climate model projections, says Zeke Hausfather, climate research lead at the payment services company Stripe and lead author of the commentary. "We must move away from the naïve idea of model democracy." Instead, he and his colleagues call for a model meritocracy, prioritizing, at times, results from models known to have more realistic warming rates.

Overall, climate models remain incredibly successful research tools, and nothing about this "too hot" generation invalidates the tenets of climate science, says Kate Marvel, a climate scientist at NASA's Goddard Institute for Space Studies and co-author of the commentary. The greenhouse effect is still warming the planet. Ice is melting, seas are rising, and droughts are becoming more frequent in some areas. But the models are not perfect, Marvel says. "They're not crystal balls."

The problem of the too-hot models arose in 2019 from the Coupled Model Intercomparison Project (CMIP), which combines the results of the world's models in advance of the major reports that come out every 7 or 8 years from the Intergovernmental Panel on Climate Change (IPCC). In previous rounds of CMIP, most models projected a "climate sensitivity"—the warming expected when atmospheric carbon dioxide is doubled over preindustrial times—of between 2°C and 4.5°C. But for the 2019 CMIP6 round, 10 out of 55 of the models had sensitivities higher than 5°C—a stark departure. The results were also at odds with a landmark study that eschewed global modeling results and instead relied on paleoclimate and observational records to identify Earth's climate sensitivity. It found that the likely value sits somewhere between 2.6°C and 3.9°C.

Researchers have since tracked down the causes of the too-hot models, which include those produced by the National Center for Atmospheric Research, the U.S. Department of Energy, the United Kingdom's Met Office, and Environment and Climate Change Canada. They often relate to the way models render clouds; one result has been excessive predicted warming in the tropics.

IPCC tried to compensate for this problem last year when it published its first working group report, which covers the physical basis of climate change. IPCC rated models on their skill at capturing past historical temperatures. Then, it used the best-performing models to produce its official "assessed warming" projections for different fossil fuel emissions scenarios. When it came to studying the future changes to Earth, IPCC reported results from all the models based on degree of warming: 1.5°C, 2°C, 3°C. That kept useful information from the hot models, such as changes in rainfall patterns, even if they reached the warming thresholds too fast.

Although IPCC rose to the challenge, it didn't do a great job telling everyone about the actual problem, says Hausfather, himself an IPCC co-author. "A large number of our colleagues had no idea that the IPCC did this," he says. And since then, dozens of published studies have used projections based on the raw average of all CMIP6 models or, worse, the output of individual "hot" models. "It's not because anybody is acting in bad faith," Marvel says. "It's just because there's no guidance."

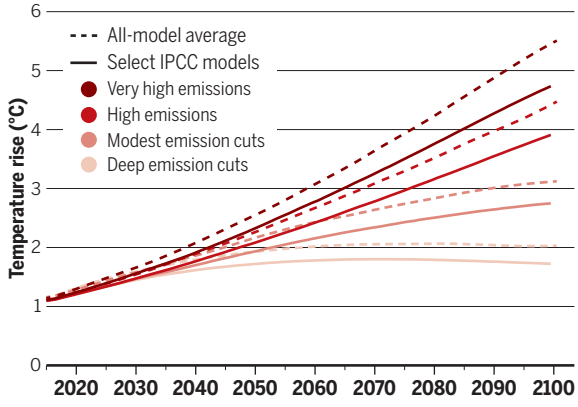
Climate impact researchers need to emulate the steps IPCC took, Hausfather and his co-authors say. First, they should avoid the dubious time-based scenarios and instead emphasize the effects of specific levels of global warming, regardless of the date those levels are reached. They should also use IPCC's assessed projections for when those warming levels might arise. And for studies where the details of the warming trajectory is important, they can use select models that predict it better, like those produced by NASA and the National Oceanic and Atmospheric Administration.

Claudia Tebaldi, a climate scientist at Lawrence Berkeley National Laboratory, agrees with most of the group's recommendations. However, she says, they may underestimate policymakers' desire for time-based information, which in her experience is nearly always requested. And some climate impacts, like sea level rise, depend on the warming rate, not just the absolute amount of warming.

Researchers should think about going even further, and examining whether certain models have, for example, big regional biases, says Reto Knutti, a climate scientist at ETH Zürich who has called for "model meritocracy" for more than a decade. As more city planners and outside scientists turn to these projections, they should first be sure to consult a climate model expert. "Given that these results guide climate adaptation and investments of billions of dollars, that seems like an effort worth doing," Knutti says. ■

Hot mess

Because of problems rendering clouds, some climate models run hot. Researchers say it's better to use select models than an average.



GRAPHIC: N. DESAI/*SCIENCE*

THE BIAS HUNTER

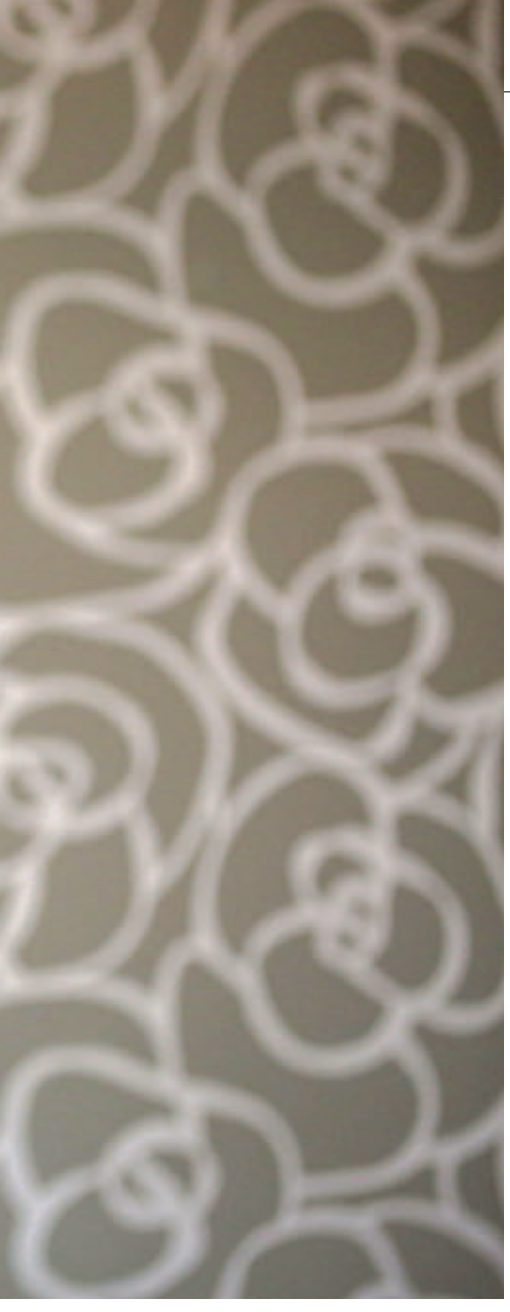
Itiel Dror is determined to reveal the role of bias in forensics, even if it sparks outrage

By **Douglas Starr**

In February 2021, cognitive psychologist Itiel Dror set off a firestorm in the forensics community. In a paper, he suggested forensic pathologists were more likely to pronounce a child's death a murder versus an accident if the victim was Black and brought to the hospital by the mother's boyfriend than if they were white and brought in by the grandmother. It was the latest of Dror's many experiments suggesting forensic scientists are subconsciously influenced by cognitive biases—biases that can put innocent people in jail.

Dror, a researcher at University College London (UCL), has spent decades using real-world cases and data to show how experts in fields as diverse as hospital care and aviation can reverse themselves when presented with the same evidence in different contexts. But his most public work has involved forensic science, a field reckoning with a history of unscientific methods. In 2009, the National Research Council published a groundbreaking report that most

PHOTO: SUSANNAH IRELAND



forensic sciences—including the analysis of bullets, hair, bite marks, and even fingerprints—are based more on tradition than on quantifiable science. Since then, hundreds of studies and legal cases have revealed flaws in forensic sciences.

Dror's work forms a connective tissue among them. He has shown that most problems with forensics do not originate with "bad apple" technicians who have infiltrated crime labs. Rather they come from the same kind of subconscious bias that affects everyone's daily decisions—the shortcuts and generalizations our brains rely on to process reality. "We don't actually see the environment," Dror says. "We perceive stimuli from the environment that our brain represents to us," shaped by feelings and past experience.

"In the span of a decade, cognitive bias went from being almost totally unheard of in forensics to common knowledge in the lab," Brandon Garrett, a professor at the Duke University School of Law, wrote in his book *Autopsy of a Crime Lab: Exposing the Flaws in Forensics*. "We can especially thank Itiel Dror for helping bring about the sea change."

Dror now travels the world testifying in trials, taking part in commissions, and offering training to police departments, forensic laboratories, judges, militaries, corporations, government agencies, and hospitals. National agencies, forensic labs, and police forces have adopted his approach to shielding experts from information that could bias them.

"I don't know anybody else who's doing everything that Itiel is doing," says Bridget Mary McCormack, chief justice of the Michigan Supreme Court, who worked with Dror on a U.S. Department of Justice task force and collaborated on studies with him. "His work is monumentally important to figuring out how we can do better. To my mind it's critical to the future of the rule of law."

"Fingerprints don't lie. ... but it's also true that fingerprints don't speak. It's the human examiner who makes the judgment, and humans are fallible."

Itiel Dror, University College London

Dror's previous studies on bias in forensics caused grumbling, but nothing like the reaction to the 2021 paper. This time, he used a survey to see whether bias could affect decision-making among medical examiners. He concluded that nonmedical evidence such as the race of the decedent or their relation to the caregiver—details that most medical examiners routinely consider—were actually a source of bias.

Eighty-five of the country's most prominent pathologists demanded its retraction. The National Association of Medical Examiners (NAME) alleged ethical misconduct and demanded that Dror's employer, UCL, stop his research. The editor of the *Journal of Forensic Sciences* wrote that he hadn't seen so many arguments in the journal's 65-year history, or so much anger. After decades challenging forensic experts, Dror had gotten into a fight that threatened his career.

DROR APPEARS to be a mild-mannered man, with salt-and-pepper hair and wire-rimmed glasses; but that impression disappears the moment he begins talking. He gains momentum like a runaway train, detailing his latest study, making a quick detour to bring up an example, slipping in a funny anecdote, then circling around to put a cap on his original point. He speaks with a mixture of accents and intonations from his upbringing in Israel, his graduate work in the United States, and his professional life in the United Kingdom.

Two diamond studs in his left ear hint at a nonconformist streak. As the child of academic parents who took frequent sabbaticals, Dror attended five elementary schools on three continents. "I'd be the new kid who didn't know the language very well," he says. "I didn't have time to assimilate or conform. It was very difficult, but it gave me a lot of independence of thought."

Dror hated reading and the discipline of school. But things turned around at age 19, after he broke his back during paratrooper training in the Israeli army—an incident he describes as a "destructively creative" shake-up to his system. Confined to a body cast for 7 months, he took to reading the books he had ignored during high school. He home-tested, and got As in the courses he had previously failed. He studied philosophy at Tel Aviv University, took a year off to work on a kibbutz, and later did a doctorate in psychology at Harvard University, studying mental imagery and decision-making.

One of his projects examined how U.S. Air Force pilots use mental imagery to recognize enemy jets traveling at high speeds. The work caught the attention of David Charlton, a prominent British fingerprint examiner who had started to have doubts about his field.

"I often wondered if when making fingerprint comparisons my eyes were the same from one day to the next," Charlton says. "And then I came across this paper suggesting that the perception of aircraft pilots could change, depending on stresses or circumstances. And I wondered if it applies to fingerprints as well."

He had reason for concern. The United Kingdom had been shaken by the scandal of Shirley McKie, a Scottish police constable who was charged with perjury after investigators claimed to find her thumbprint at a murder scene in 1997. McKie was cleared when two American experts testified that the thumbprint could not have been hers. The Americans had their own scandal in 2004, when FBI detained an American lawyer, Brandon Mayfield, as a suspect in a terrorist bombing of a Madrid train station. Among 20 near-matches

in their fingerprint database, agents focused on Mayfield, who had converted to Islam and provided legal defense to a Portland, Oregon, resident with Taliban connections. When Spanish authorities found the real bomber, Mayfield sued the U.S. government, which agreed to a $2 million settlement.

Those cases deepened Charlton's doubts about his own objectivity. He contacted Dror, who suggested they do some research together. They found five fingerprint experts who knew about the Mayfield case but had not seen the fingerprints. Dror and Charlton sent each expert a pair of prints from one of the expert's own previous cases, which they had personally verified as "matched," but told them the prints came from the notorious case of FBI's mismatch of Mayfield's prints with the terrorist's.

Four of the five experts contradicted their previous decision: Three now concluded the pair was a mismatch, and one felt he needed more information. They seemed to have been influenced by the passage of time and extraneous information.

"It was so simple and elegant," Peter Neufeld, co-founder of the Innocence Project, says of the study. "And when people in the forensic community read it, they got it."

In a follow-up study, Dror and Charlton gave six experts sets of prints they had previously examined along with biasing information—that the suspect had either confessed or had an alibi. Four of the six experts changed their past findings.

The results turned some of Charlton's colleagues against him. "A lot of people wondered if I was trying to destroy the profession," he says. Angry letters poured in to *Fingerprint Whorld*, the professional journal of which Charlton was editor. The chair of the Fingerprint Society wrote that any fingerprint examiner who could be swayed by images or stories "is so immature he/she should seek employment at Disneyland."

Charlton was so upset by these reactions that he considered abandoning his career. "Don't worry, this is normal," he remembers Dror telling him. "It's part of the human condition. Now let's do more research and see how we can improve things."

Dror looked at other biasing factors in fingerprint analysis, some of which were shockingly innocuous. When police retrieve a print from a crime scene, they consult an FBI computer database containing millions of fingerprints and receive several possible matches, in order of the most likely possibilities. Dror found that experts were likely to pick "matches" near the top of the list even after he had scrambled their order, perhaps because of the subconscious tendency to overly trust computer technology.

Kerry Robinson was exonerated after serving 18 years for a conviction based in part on contested DNA analysis.

"People would say to me fingerprints don't lie," Dror says. "And I would say yes, but it's also true that fingerprints don't speak. It's the human examiner who makes the judgment, and humans are fallible."

Dror and his colleagues are quick to point out that bias does not always equal prejudice, but it can foster injustice. Studies have shown, for example, that Black schoolchildren get punished more readily than white children for the same misbehavior, because many teachers subconsciously assume Black children will continue to misbehave. And in forensic science, bias can subconsciously influence experts to interpret data in a way that incriminates a suspect.

If something as seemingly infallible as fingerprints could be biased, what could be next? Dror set his sights on DNA. When the authors of the National Research Council study criticized forensic sciences, they made an exception for DNA analysis, a method developed in the lab that was statistically verifiable and scientifically sound.

But as DNA analysis has gotten more sensitive and sophisticated, it has also come to rely more on human interpretation. For example, when investigators find a mixture of several people's DNA at a crime scene, it's up to the analyst to tease apart the contributors. It's a complicated and subtle process, one that Dror found can be influenced by context. Consider the case of Kerry Robinson in Georgia, who was accused in 2002 of taking part in a gang rape. The state based its case on the plea bargain testimony of Tyrone White, who investigators had identified as the main perpetrator and who bore Robinson a grudge. The state's two DNA experts found that Robinson's DNA "could not be excluded," from the mixture of DNA found at the crime scene, and the jury found him guilty.

Greg Hampikian, a genetics professor then at the Georgia Innocence Project, sent DNA data from the case to Dror, who shared it with 17 DNA analysts unfamiliar with the case. Only one agreed with Georgia's analysts; the other 16 either excluded Robinson's DNA or said they could not come up with a result. Dror's conclusion: Even DNA analysis, the "gold standard" of forensic science, was subject to human bias. The state did not release Robinson until 2020, when Hampikian submitted other exonerating information. Robinson had already served 18 years of his 20-year sentence.

PHOTO: GEORGIA INNOCENCE PROJECT

Over the years Dror and other researchers have found bias just about everywhere they've looked—in toxicologists, forensic anthropologists, arson investigators, and others who must make judgments about often ambiguous crime scene evidence. Yet juries find forensic evidence compelling, Dror and others have found.

Many examiners feel "impervious to bias," says Saul Kassin, a psychologist at John Jay College of Criminal Justice, "as if they're not human like the rest of us." In 2017, Kassin and Dror asked more than 400 forensic scientists from 21 countries about their perceptions of bias. They found that whereas nearly three-quarters of the examiners saw bias as a general problem, just over 52% saw it as a concern in their own specialty, and only 26% felt that bias might affect them personally.

Dror says the best approach to fighting bias is to shield experts from extraneous information, similar to the "blinding" in scientific experiments. He calls the process Linear Sequential Unmasking, in which the analyst only sees the evidence that's directly relevant to their task. Some authorities have endorsed the approach. The United Kingdom's Forensic Science Regulator recommends it as "the most powerful means of safeguarding against the introduction of contextual bias." FBI adopted the process following the Mayfield case: Because humans tend to see similarities between objects viewed side by side, agents now document the features of a crime scene fingerprint on its own before comparing it to a suspect's prints.

After consultation with Dror, police in the Netherlands began to blind fingerprint examiners to details of a crime investigation that might influence their analysis, such as the condition of the body or the urgency of the case, says John Riemen, the police force's lead biometrics specialist. The approach ensures "you're looking at fingerprints, and not at your biases," he says.

IT WAS AN ATTEMPT to win medical examiners over to this approach that landed Dror in hot water. In 2019, he got a message from Daniel Atherton, a pathologist at the University of Alabama, Birmingham, who wanted him to look at some data he had collected. Atherton had sent a survey to 713 pathologists across the country positing one of two scenarios in which a toddler with a skull fracture and brain hemorrhage was brought to an emergency room and died shortly thereafter. In one scenario, the child was white and was brought in by the grandmother. In the other, the child was Black and brought in by the mother's boyfriend. The survey asked participants to decide whether the manner of death was undetermined, accidental, or homicide.

Dror analyzed the results and found that of the 133 people who answered the survey, 32 concluded the death was a homicide. And a disproportionate number of those—23—had received the scenario with the Black child and the boyfriend. Participants reading the "Black condition" were five times more likely to conclude homicide than accident, whereas participants in the "White condition" ruled accident more than twice as frequently as homicide.

"Their decisions were noticeably affected by medically irrelevant contextual information," Dror, Atherton, and their colleagues wrote in their paper, published in the *Journal of Forensic Sciences*.

The paper also included a survey of 10 years of Nevada death certificates showing an apparent correlation between Black deaths and findings of homicide versus accident—influenced, perhaps, by cultural biases. "I just wanted to get that information out there to begin a discussion," Dror says of the study.

He got more of a discussion than he expected. The journal was swamped with angry letters from medical examiners. One derided the study as "rank pseudoscience." Another, signed by the president of NAME along with 84 other pathologists, excoriated the study as "fatally flawed" and "an abject failure of the peer review process," and demanded its retraction. (Michael Peat, editor of the journal, declined to retract the article, saying it had been peer reviewed before publication and rereviewed by a respected biostatistician following the complaints.)

Many pathologists pointed out that the experimental design linked two unrelated variables—the race of the child and their relationship to the caretaker. They were further inflamed by Dror's labeling the scenarios "Black condition" and "White condition," when they had reason to suspect that the caretaker, not the race, was the relevant variable. Statistics show a boyfriend of any race is far more likely to harm a child in his care than a grandmother.

"To introduce race … appears to be an effort to label the survey responders, and their colleagues by proxy, as racist," said the letter from the 85 practitioners. "Had this survey been done with the races reversed … White cases were more likely to be called homicide and Black cases more likely to be called accident." They contended that Dror was using inflammatory language to get headlines. And they noted that other factors could have played a role in the pathologists' decisions, such as their level of experience, local crime statistics, and office policies, none of which Dror had considered.

Stephen Soumerai, an expert in research design at Harvard Medical School, agrees that linking a known risk factor for homicide (caregiver relationship) to a nonwhite race is problematic. And the survey of Nevada death certificates failed to investigate other possible explanations beyond race, he says. "The hypothesis is reasonable and important, but the research does not adhere to basic principles of research design," he says.

Dror admits he would have been wise to use neutral terms to designate the two experimental groups. But he doesn't concede that the study is flawed. "It is a first study to examine and establish that there is bias

A glossary of bias

Itiel Dror and his collaborators have coined various terms to describe how bias sneaks into forensic analysis—and how experts perceive and react to their biases.

Target-driven bias	Subconsciously working backward from a suspect to crime scene evidence, and thus fitting the evidence to the suspect—akin to shooting an arrow at a target and drawing a bull's-eye around where it hits
Confirmation bias	Focusing on one suspect and highlighting the evidence that supports their guilt, while ignoring or dismissing evidence to the contrary
Bias cascade	When bias spills from one part of the investigation to another, such as when the same person who collects evidence from a crime scene later does the laboratory analysis and is influenced by the emotional impact of the crime scene
Bias snowball	A kind of echo chamber effect in which bias gets amplified because those who become biased then bias others, and so on
Bias blind spot	The belief that although other experts are subject to bias, you certainly are not
Expert immunity	The belief that being an expert makes a person objective and unaffected by bias
Illusion of control	The belief that when an expert is aware of bias, they can overcome it by a sheer act of will
Bad apples	The belief that bias is a matter of incompetence or bad character
Technological protection	The belief that the use of technology, such as computerized fingerprint matching or artificial intelligence, guards against bias

in forensic pathology," he says. Dror agrees that statistics do show an unrelated caretaker is more likely to harm a child than a grandmother. But such generalizations should not affect how examiners diagnose individual cases.

Judy Melinek, CEO of PathologyExpert, Inc. who practices forensic pathology in Wellington, New Zealand, agrees. "I've seen too many cases where innocent caregivers were prosecuted for accidental child deaths because forensic pathologists made assumptions based on larger trends."

Dror says antibias strategies are especially important to medical examiners because many work hand in hand with police who might influence them. One solution is to have a laboratory's case manager safeguard details about an investigation and unveil them to a medical examiner only as needed to determine the manner of death—similar to the Linear Sequential Unmasking used by the Dutch police, among others. "It's all about looking at the right evidence in the right sequence," Dror says.

Brandon Mayfield was detained as a bombing suspect based on a flawed FBI fingerprint analysis.

That approach represents a "clueless" understanding of how medical examiners work—one that cognitive psychologists have held for years, says William Oliver, a retired professor of pathology at East Carolina University's Brody School of Medicine and a former board member of NAME. Unlike other forensic examiners, who match patterns from a particular type of evidence, medical examiners must gather all the information about the case that they can to make a correct diagnosis, he says.

They determine both cause of death—the injury or illness that killed a person—and the manner of death, which describes how the death came about. If a dead man is found sitting in his car with the engine idling, the garage door closed, and high levels of carbon monoxide in his blood, the autopsy would likely conclude that the cause of death was carbon monoxide poisoning. But the manner of death would remain "undetermined" unless investigators found signs pointing to suicide, such as a note, recent job loss or divorce, or statements from friends that he had been depressed.

"Manner is not a scientific determination, and it is not meant to be," Oliver says. Aggregate statistics—like the rates at which grandmothers and unrelated caretakers harm children—are crucial to making that judgment, he says.

The acrimony around Dror's paper snowballed. On 19 March 2021, Brian Peterson, a member of NAME and chief medical examiner for Milwaukee County in Wisconsin, filed a formal ethics complaint with the association against the four pathologists who collaborated with Dror. Their paper would "do incalculable damage to our profession," exposing every medical examiner to withering cross-examination at trials, he wrote.

"I was shocked at the reaction," says Joye Carter, a forensic pathology consultant and co-author on the study, who was named in the complaint. "We're supposed to be fact finders, but people got whipped up into this ridiculous attitude that they were being persecuted."

Both the Innocence Project and the Legal Defense Fund came to the pathologists' defense, and NAME dismissed Peterson's complaint in May 2021. (Carter, a prominent pathologist who was the first Black chief medical examiner in the United States, resigned from NAME. "There's no way I can be part of a group like this," she says.)

But Dror faced a separate attack from NAME's leadership. In an 8 March 2021 letter to UCL, NAME's then-President James Gill and Executive Vice President Mary Ann Sens accused him of intentional ethics violations, including misleading participants by not telling them the study was about race and bias.

The letter triggered a hearing at UCL's ethics board, which Dror says could have ended in his dismissal. He argued that disclosing the nature of the study would have biased the results, and at one point he became so emotional that he had to leave the room to regain his composure. Ultimately, the board found in his favor, ruling that "the allegation is mistaken."

The question of bias in autopsies rocketed to the headlines after Minneapolis police officer Derek Chauvin killed George Floyd on 25 May 2020. During the trial in April 2021, the local medical examiner for Hennepin County in Minnesota testified that the manner of death was "homicide," as did other pathologists. But an expert hired by Chauvin's defense team, former Maryland Chief Medical Examiner David Fowler, testified that Floyd had so many underlying health challenges that the manner of death was "undetermined."

Chauvin was found guilty, but Fowler's testimony outraged other pathologists and physicians, who saw in his conclusions a pro-police bias. More than 400 of them signed a petition to Maryland Attorney General Brian Frosh demanding an investigation into all the death-in-police-custody cases during Fowler's 17 years in office. Frosh recruited seven international experts to design the study, including Dror. And despite all the blowback Dror has received for trespassing in the field of forensic patho-logy, he agreed to participate.

"If my work results even in one person not getting wrongly convicted, or one guilty person not going free, then it's worth all the grief I've been getting," he says. "And maybe not just one person. Hopefully this is going to change the domain." ■

Douglas Starr is a journalist in the Boston area.

PHOTO: AP PHOTO/DON RYAN

eppendorf & Science

PRIZE FOR NEURO BIOLOGY

2021 Winner
Amber L. Alhadeff, Ph.D.
Monell Chemical Senses Center, USA
For research on the gut-brain control of hunger circuits

Tell the World About Your Work!

Application Deadline
June 15, 2022

Eppendorf & Science Prize for Neurobiology
The annual Eppendorf & Science Prize for Neurobiology is an international prize which honors young scientists for their outstanding contributions to neurobiological research based on methods of molecular and cell biology. The winner and finalists are selected by a committee of independent scientists, chaired by *Science's* Senior Editor, Dr. Peter Stern. If you are 35 years of age or younger and doing great research, now is the time to apply for this prize.

As the Grand Prize Winner, you could be next to receive
> Prize money of US$25,000
> Publication of your work in *Science*
> Full support to attend the Prize Ceremony held in conjunction with the Annual Meeting of the Society for Neuroscience in the USA
> 10-year AAAS membership and online subscription to *Science*
> Complimentary products worth US$1,000 from Eppendorf
> An invitation to visit Eppendorf in Hamburg, Germany

It's easy to apply! Write a 1,000-word essay and tell the world about your work. Learn more at:

eppendorf.com/prize

eppendorf Science AAAS

AAAS® and Science® are registered trademarks of the American Association for the Advancement of Science, USA. Eppendorf® and the Eppendorf Brand Design are registered trademarks of Eppendorf SE, Germany. All rights reserved, including graphics and images. Copyright © 2022 by Eppendorf SE. Photography: Happy Hour Headshot Philadelphia.

INSIGHTS

PERSPECTIVES

CLIMATE CHANGE

The limits of forest carbon sequestration

Current models may be overestimating the sequestration potential of forests

By **Julia K. Green**[1] *and* **Trevor F. Keenan**[1,2]

Although nations are grappling with reducing carbon emissions to minimize climate change risks, they actually need to remove carbon dioxide from the atmosphere if the goal is to keep warming to less than 1.5°C (*1*). Although there has been substantial investment in technologies focusing on carbon capture, there is also interest in natural solutions, such as planting trees (*2*, *3*). Trees remove CO_2 from the atmosphere through photosynthesis, and forests sequester CO_2 in the form of biomass and soil carbon. Though it seems logical that increased photosynthesis should lead to increased growth, multiple reports suggest that this is often not the case (*4*, *5*). The interaction between photosynthesis, tree growth, and carbon sequestration has large implications for future carbon uptake and is a topic that remains widely debated. On page 758 of this issue, Cabon *et al.* (*6*) contribute to this debate by examining global photosynthesis estimates and tree ring measurements as a measure of growth.

Carbon sequestration in land ecosystems has been gradually increasing since the 1970s (*7*). CO_2 is a necessary component for plant photosynthesis, and rising concentrations of atmospheric CO_2 from anthropogenic emissions have led to higher rates of carbon uptake by plants, a phenomenon known as carbon fertilization (*8*). Additionally, increases in greenhouse gases have lengthened the growing season indirectly through warming, leading to additional increases

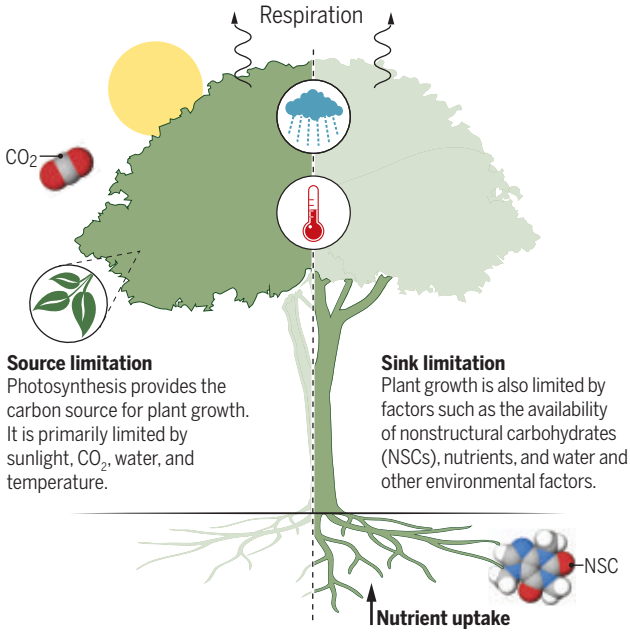
Cabon *et al*. highlight the importance of considering processes other than photosynthesis to estimate how much carbon trees can sequester.

PHOTO: VOISIN/PHANIE/SCIENCE SOURCE

in ecosystem carbon uptake (*7*, *9*, *10*). These increases in response to rising atmospheric CO_2 concentration have led to claims that ecosystems are mostly limited by the amount of atmospheric CO_2. However, a large body of research has indicated that increases in atmospheric CO_2 do not necessarily translate to increased carbon sequestration because of other strong limitations on plant growth (*11*–*13*). Even if more carbon is available to vegetation, other factors are needed to help stoke growth from photosynthesis. Most notably, nutrient abundance and water content in the soil, as well as temperature, can limit growth (see the figure). This complication, best characterized as a dichotomy of whether forest carbon sequestration is carbon source limited (limited by how much carbon the plant can photosynthesize) or sink limited (limited by how much the plant can grow because of all environmental factors), has large implications for how much forests can sequester carbon.

Source and sink limitations on plant growth

Plant growth can be restricted by source and/or sink limitations. Although source and sink limitations share several controlling factors, the time scales on which they operate and their sensitivity to these factors often differ.


Respiration
CO_2
Source limitation
Photosynthesis provides the carbon source for plant growth. It is primarily limited by sunlight, CO_2, water, and temperature.
Sink limitation
Plant growth is also limited by factors such as the availability of nonstructural carbohydrates (NSCs), nutrients, and water and other environmental factors.
NSC
Nutrient uptake

To understand the restrictions on plant growth related to source and sink limitations, Cabon *et al.* used gross primary production (GPP) estimates—which denote the amount of carbon taken up by plants during photosynthesis—and combined them with tree ring data from the International Tree-Ring Data Bank to perform a cross-biome correlation analysis. They compiled the most comprehensive tree ring dataset to date and identified source-sink relationships across the globe. They demonstrated that there is a strong decoupling between vegetation photosynthesis and the radial growth of a tree trunk, which shows substantial variation based on vegetation type (e.g., angiosperm versus gymnosperm), ecosystem traits (e.g., forest age and nutrient availability), and climate (e.g., more decoupling in cooler, more arid climates). These results emphasize that the relationship between photosynthesis and growth is not as simplistic as the current representation in models, which typically depict vegetation carbon sequestration as being closely related to GPP (*14*).

The results reported by Cabon *et al.* have implications for using natural ecosystems to sequester carbon and for the success of natural climate solutions, such as planting trees, in combating climate change. If less carbon than predicted will be stored in biomass in the coming years, the utility of this ecosystem service is greatly diminished. Additionally, the results emphasize caution when interpreting findings solely on the basis of GPP estimates that make generalizations about carbon storage. For example, a GPP increase in the springtime from plants producing their leaves earlier may not translate to an increased overall growth by the end of the year.

Nevertheless, the use of GPP estimates and tree ring data to understand the degree to which growth and carbon sequestration by trees are inherently sink limited does have some constraints. For example, related to the use of GPP data, a portion of the CO_2 absorbed during photosynthesis is released back into the atmosphere through plant respiration. Therefore, the amount of carbon available to plants for growth is actually not GPP, but the carbon gains during GPP minus the autotrophic respiration losses—i.e., the net primary production (NPP). Many studies examining the relationship between plant carbon uptake and growth have used NPP data, which are more directly linked to vegetation growth (*15*) and have a stronger relationship with biomass. GPP is therefore not necessarily a reliable proxy for the carbon source available for growth. For this reason, studies that measured NPP at some of the same sites as Cabon *et al.* have reported a stronger correlation between carbon source activity and tree ring widths, without a role of sink limitation (*15*). Also, related to the use of tree ring data to infer tree growth, plants allocate carbon not only to woody aboveground biomass, but also to their root systems, their leaves, and pools of nonstructural carbohydrates that are stored throughout the plant. Plants alter their carbon allocation depending on environmental changes, forest age, and the time of year. Therefore, there are uncertainties when one infers the amount of carbon sequestered based solely on tree ring measurements because woody growth is not synonymous with total carbon sequestered.

Despite these limitations, Cabon *et al.* show that carbon projections using the current modeling framework may be overestimating the carbon sequestration in forests. Their findings highlight the need for an in-depth assessment of the degree to which vegetation models effectively represent growth dynamics and the degree of carbon sink and source limitations. To do so will require a better understanding of these limitations, which should be possible with new methodologies of measuring autotrophic respiration and NPP directly at flux tower sites, as well as an increase in the number of collocated measurements related to carbon allocation and sequestration. ■

[1]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA. [2]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. Email: jgreen17@berkeley.edu

GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2022: Impacts, Adaptation, and Vulnerability*, Contribution of Working Group II to the Sixth Assessment Report of the IPCC (Cambridge Univ. Press, 2022).
2. P. H. S. Brancalion, K. D. Holl, *J. Appl. Ecol.* **57**, 2349 (2020).
3. G. M. Domke, S. N. Oswalt, B. F. Walters, R. S. Morin, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24649 (2020).
4. N. Delpierre, D. Berveiller, E. Granda, E. Dufrêne, *New Phytol.* **210**, 459 (2016).
5. R. L. Peters *et al.*, *New Phytol.* **229**, 213 (2021).
6. A. Cabon *et al.*, *Science* **376**, 758 (2022).
7. P. Friedlingstein *et al.*, *Earth Syst. Sci. Data* **12**, 3269 (2020).
8. A. P. Walker *et al.*, *New Phytol.* **229**, 2413 (2021).
9. T. F. Keenan *et al.*, *Nat. Commun.* **7**, 13428 (2016).
10. D. Schimel, B. B. Stephens, J. B. Fisher, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 436 (2015).
11. S. Fatichi *et al.*, *New Phytol.* **221**, 652 (2019).
12. S. Fatichi, S. Leuzinger, C. Körner, *New Phytol.* **201**, 1086 (2014).
13. M. Jiang *et al.*, *Nature* **580**, 227 (2020).
14. A. D. Friend *et al.*, *Ann. For. Sci.* **76**, 49 (2019).
15. A. D. Richardson *et al.*, *New Phytol.* **197**, 850 (2013).

10.1126/science.abo6547

AGING

Immune cells impede repair of old neurons

Interfering with age-related neuroimmune interactions promotes nerve regeneration

By **Nicolas Gaudenzio**[1,2] *and* **Roland S. Liblau**[1,3]

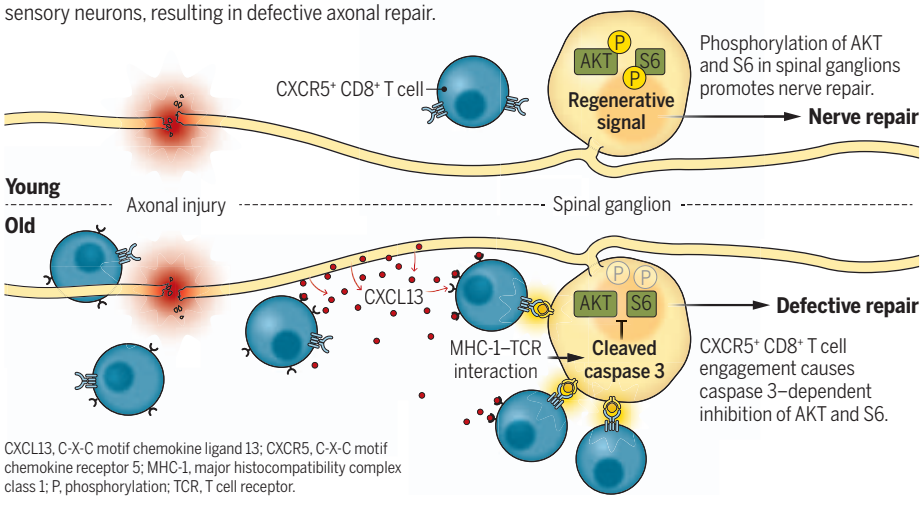
The regenerative capacity of older people is reduced, resulting in decreased tissue function and resilience. Accordingly, the regeneration of the sciatic nerve after injury has been reported to be less efficient and slower in older people (*1*). One of the hallmarks of aging is altered intercellular communication, which is often accompanied by increased density of immune cells within tissues and excessive release of proinflammatory mediators, called inflammaging (*2, 3*). In this context, the immune system disturbs tissue homeostasis and impedes functional recovery. However, the precise mechanisms underlying this pathophysiological process are largely elusive, which is a barrier to rational treatment design. On page 715 of this issue, Zhou *et al.* (*4*) describe a mechanism by which aged sensory neurons release the chemoattractive protein C-X-C motif chemokine ligand 13 (CXCL13). Upon sciatic nerve injury in aged, but not young, mice, this results in the recruitment of CD8$^+$ T cells that prevent axonal regeneration.

T cells are endowed with both antigen specificity and memory capacities, and neuron–T cell interactions have been largely documented during infectious, degenerative, and autoimmune diseases of the nervous system (*5–7*). Neurons can play a central role in this process by releasing inflammatory mediators and recruiting immune cells to the site of tissue damage, thereby contributing to usually beneficial and reparative inflammatory processes. The development of a local inflammatory reaction needs to be finely regulated to avoid the destruction of postmitotic cells such as neurons. However, the nervous system (and many other tissues) is gradually infiltrated by antigen-experienced T cells upon aging. This progression is well documented by Zhou *et al.* in the peripheral nervous system and by recent reports in the brain and spinal cord in both mice and humans (*8–11*). The increased accessibility of aged T cells to distant tissues likely represents an important contributing factor (*2*, *8*). Moreover, repeated tissue insults and infections, which accumulate with age, may each leave an additional local population of T cells that can rapidly fight the previously encountered intruder. Of note, the increased density of CD8$^+$ T cells with cytotoxic potential may play a beneficial role through the elimination of senescent cells. The proportion of senescent cells increases in tissues with age; they undergo a stable cell-cycle arrest and secrete proinflammatory and profibrotic factors and thereby interfere with organ function and contribute to age-related diseases (*12*).

[1]Toulouse Institute for Infectious and Inflammatory Diseases (Infinity), Inserm UMR1291 CNRS UMR5051, University of Toulouse III, Toulouse, France. [2]Genoskin SAS, Toulouse, France. [3]Department of Immunology, Toulouse University Hospital, Toulouse, France. Email: roland.liblau@inserm.fr

Age-associated nerve injury responses

In young mice, nerve injury usually triggers a regenerative program, which leads to nerve repair. In old mice, there is a high density of CD8$^+$ T cells, and injury causes the release of CXCL13 by sensory neurons. This recruits CXCR5$^+$ CD8$^+$ T cells, which engage MHC-1–antigen complexes expressed on the cell body of injured sensory neurons, resulting in defective axonal repair.



CXCL13, C-X-C motif chemokine ligand 13; CXCR5, C-X-C motif chemokine receptor 5; MHC-1, major histocompatibility complex class 1; P, phosphorylation; TCR, T cell receptor.

However, recent reports suggest that the protection afforded by tissue-resident CD8$^+$ T cells can be hijacked to favor age-associated miscalibrated responses or repair decline. Consistently, age-associated brain CD8$^+$ T cells exacerbate acute ischemic injury and neurodegeneration following traumatic brain injury (*8*). Increased numbers of brain-infiltrating CD8$^+$ T cells have also been documented in neurodegenerative diseases, such as Alzheimer's disease, although their precise role is unknown (*13*). Zhou *et al.* unveil a previously unrecognized molecular communication between sensory neurons and immune cells that is inherent to the aged nervous system, and which is triggered by an aberrant release of neuron-derived CXCL13, as a result of increased expression of the phosphorylated form of the transcription factor nuclear factor κB2 (NF-κB2). Upon sciatic nerve injury in aged, but not young, mice, the release of CXCL13 promotes the recruitment of CD8$^+$ T cells expressing CXCR5 (the receptor for CXCL13) to the spinal ganglion where the cell bodies of sensory neurons are located. These recruited CD8$^+$ T cells interact physically with sensory neuron cell bodies and prevent axonal regeneration, through a caspase 3–dependent reduction in phosphorylated AKT and S6 ribosomal protein. Consequently, nerve repair is defective in old animals (see the figure).

Albeit of different focus, a recent study revealed the accumulation of antigen-experienced CD8$^+$ T cells that produce the cytokine interferon-γ in old mouse brains that can alter the proliferation of neural stem cells in the subventricular zone (*9*). Such cytokine-mediated effects could erode regenerative capacities and thereby contribute to brain aging, possibly reflecting another facet of the process identified by Zhou *et al.*

During aging, transmigration of T cells across tissues increases, in part owing to greater permeability of blood vessels (*14*), which likely contributes to their accumulation in tissues over time. Intriguingly, senescent T cells have been reported to propagate certain features of premature aging throughout the body (*15*). Zhou *et al.* found that CD8$^+$

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

T cells infiltrating the spinal ganglion (either preexisting or attracted there upon injury) of old mice directly interact with neurons in an antigen-dependent manner, then release perforin and granzyme B that trigger neuronal caspase-3 activation without inducing cell death. The identity of the antigen(s) involved in this deleterious neuron–T cell cross-talk is unknown.

Zhou *et al.* uncover a cellular and molecular mechanism that underpins regenerative decline in the nervous system of old mice. Although it is unknown if this mechanism occurs in humans or is specific to the spinal ganglion, the authors investigated whether blocking this pathway might offer a therapeutic opportunity to promote sciatic nerve injury repair. They showed that neutralization of CXCL13 through administration of monoclonal antibodies prevents the recruitment of CXCR5-expressing $CD8^+$ T cells in the spinal ganglion, which promoted axonal regeneration and functional recovery following sciatic nerve injury in aged mice. In an elegant set of complementary experiments, Zhou *et al.* conclusively demonstrate that the specific targeting of $CD8^+$ T cells expressing CXCR5 phenocopies the beneficial effect of CXCL13 neutralization. However, CXCR5 is expressed at the cell surface of numerous cell types, including B cells and a subset of helper $CD4^+$ T cells, guiding them to lymphoid follicles where they can engage in productive bidirectional interactions. Moreover, it is important to keep in mind that CXCR5-expressing lymphocytes, including $CD8^+$ T cells, are key to mounting an appropriate immune response against viruses and likely against cancer cells. Therefore, in a future clinical setting, patients systemically and/or chronically treated with CXCL13-CXCR5 antagonists may be exposed to a higher frequency of (re)infections and/or neoplasia. Determining the best timing and route of administration of such treatment will also be key to opening a safe path for future therapeutic intervention. ■

REFERENCES AND NOTES

1. M. W. Painter *et al.*, *Neuron* **83**, 331 (2014).
2. M. Mittelbrunn, G. Kroemer, *Nat. Immunol.* **22**, 687 (2021).
3. D. A. Mogilenko *et al.*, *Immunity* **54**, 99 (2021).
4. L. Zhou *et al.*, *Science* **376**, eabd5926 (2022).
5. R. S. Liblau *et al.*, *Trends Neurosci.* **36**, 315 (2013).
6. G. Di Liberto *et al.*, *Cell* **175**, 458 (2018).
7. D. Frieser *et al.*, *Sci. Transl. Med.* **14**, eabl6157 (2022).
8. R. M. Ritzel *et al.*, *J. Immunol.* **196**, 3318 (2016).
9. B. W. Dulken *et al.*, *Nature* **571**, 205 (2019).
10. D. Mrdjen *et al.*, *Immunity* **48**, 599 (2018).
11. J. Smolders *et al.*, *Nat. Commun.* **9**, 4593 (2018).
12. Y. Ovadya *et al.*, *Nat. Commun.* **9**, 5435 (2018).
13. D. Gate *et al.*, *Nature* **577**, 399 (2020).
14. J. Amersfoort *et al.*, *Nat. Rev. Immunol.* (2022).
15. G. Desdín-Micó *et al.*, *Science* **368**, 1371 (2020).

ACKNOWLEDGMENTS

The authors are supported by the Inspire geroscience project (Occitania Region) and the European Research Council (ERC-2018-STG 802041 to N.G.).

10.1126/science.abp9878

GENOMICS

Mapping cell types across human tissues

Single-cell analyses reveal tissue-agnostic features and tissue-specific cell states

By **Zedao Liu** *and* **Zemin Zhang**

Our understanding of how cells form distinct tissues and organs and interact with each other is limited. Recent single-cell RNA sequencing (scRNA-seq) analyses have described the landscapes of individual cell types, along with their abundance and interactions, in homeostasis and diseased conditions (*1*–*5*), but these studies are often limited to a single organ. A systematic comparison of cell types across different tissues is needed to understand shared and variable transcriptional features and how these specializations are important for organ function. On pages 711, 712, and 713 of this issue, The Tabula Sapiens Consortium (*6*), Eraslan *et al.* (*7*), and Domínguez Conde *et al.* (*8*), respectively, as well as Suo *et al.* (*9*), report pan-tissue single-cell transcriptome atlases covering more than a million cells, including 500 cell types, across more than 30 human tissues from 68 donors. These four studies apply rigorous ontologies to consistently annotate and compare single cells between organs.

The interrogation of these large datasets reveals tissue-agnostic and tissue-specific cell features, identifies rare cell types, and provides insights into cell states that are likely to underlie disease pathogenesis (see the figure). The Tabula Sapiens Consortium created a human reference atlas across 24 different tissues and organs using scRNA-seq, leading to the characterization of more than 400 cell types spanning epithelial, endothelial, stromal, and immune cell compartments. Eraslan *et al.* took a complementary approach by applying single-nucleus RNA sequencing (snRNA-seq) to eight human tissue types and profiled neuronal cells, muscle cells, and adipocytes, which are hard to dissociate and capture using scRNA-seq. These cross-tissue approaches recapitulated conserved cell-type features and revealed cell-state adaptations to distinct tissue environments.

For example, the Tabula Sapiens Consortium discovered that endothelial cells from lung, heart, uterus, liver, pancreas, fat, and muscle exhibit the most distinct transcriptional signatures, suggesting highly specialized functions, whereas endothelial cells from the thymus, vasculature, prostate, and eye resemble one another. The pan-tissue approach led to the discovery of SLC14A1 (solute carrier family 14 member 1) as a marker for heart endothelial cells, likely reflecting specialized metabolism in cardiac blood vessels. Eraslan *et al.* also found rare cell types, such as neuroendocrine cells in the prostate and enteric neurons in the esophagus. Additionally, the corroborative use of both high-throughput 10X and full-length SMART-seq2 single-cell transcriptome data allowed the quantification of splicing isoform usage at the single-cell level, thereby revealing differential exon usage patterns for genes, including *MYL6* (myosin light chain 6) and *CD47*, in different cell-type compartments.

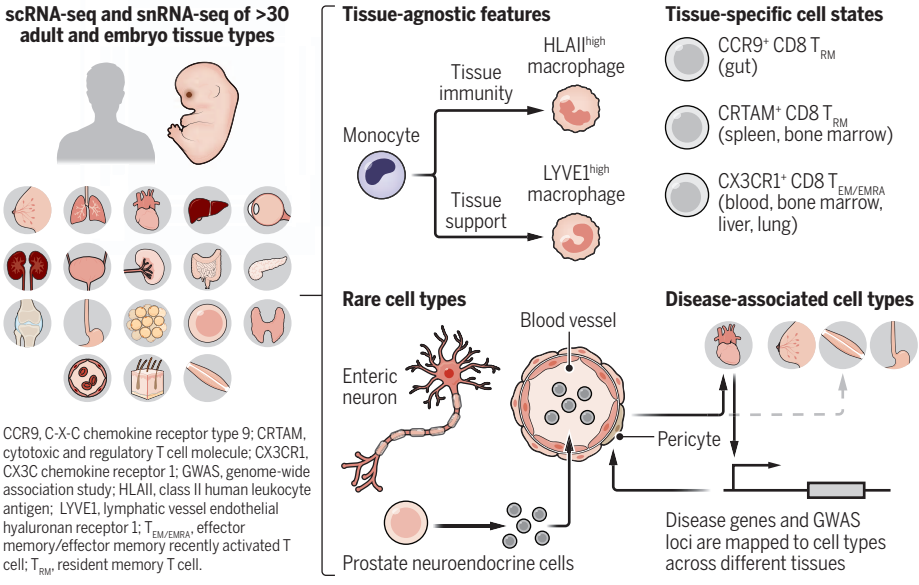
In addition to tissue-resident immobile cell types, the immune system is central to executing host defense and orchestrating tissue homeostasis. However, the current understanding of the components of the human immune system is incomplete, in part because of the multiple tissues involved in immune cell development and trafficking. Comparing immune cells from different organs could reveal shared and tissue-specific immune cell states that reflect adaptations to different local tissue niches. For example, analysis of myeloid cells across organs by Eraslan *et al.* revealed a conserved macrophage differentiation trajectory across tissues, with common monocyte precursors differentiating into macrophages that express class II human leukocyte antigen ($HLAII^{high}$) in tissue immunity and macrophages that express lymphatic vessel endothelial hyaluronan receptor 1 ($LYVE1^{high}$) for blood-vessel integrity support and inhibition of immune cell infiltration.

Domínguez Conde *et al.* and Suo *et al.* profiled human adult and developing immune cells with scRNA-seq. The adult immune cell atlas from Domínguez Conde *et al.* provides a detailed phenotyping of innate and adaptive immune cells across lymphoid and nonlymphoid organs. Tissue-specific immune cell subsets emerged, with variable chemokine and tissue-residency features reflecting their adaptation to organ microenvironments.

Biomedical Pioneering Innovation Center, Peking University, Beijing, China. Email: zemin@pku.edu.cn

Pan-tissue single-cell atlases

Four pan-tissue atlases were constructed using single-cell and single-nucleus RNA sequencing (scRNA-seq and snRNA-seq) of >1 million cells, covering 500 cell types across >30 human tissues. Cross-tissue comparison of cell types and their transcriptional features revealed rare cell types, tissue-agnostic features, and tissue-specific cell states. Disease-associated cell types were also discovered.



CCR9, C-X-C chemokine receptor type 9; CRTAM, cytotoxic and regulatory T cell molecule; CX3CR1, CX3C chemokine receptor 1; GWAS, genome-wide association study; HLAII, class II human leukocyte antigen; LYVE1, lymphatic vessel endothelial hyaluronan receptor 1; $T_{EM/EMRA}$, effector memory/effector memory recently activated T cell; T_{RM}, resident memory T cell.

Memory T cells are antigen-experienced T cells, and three subsets of memory $CD8^+$ T cells were found in the adult immune cell atlas, with effector memory/effector memory recently activated T ($T_{EM/EMRA}$) cells prevalent in blood-rich sites, resident memory T (T_{RM}) cells enriched in intestinal sites, and resident memory/effector memory ($T_{RM/EM}$) T cells predominant in spleen and bone marrow; each subset exhibited distinct chemokine receptor expression, which might drive their tissue localization.

The prenatal immune cell atlas from Suo *et al.* compares cell states across gestation stages and reveals conserved and tissue-specific trajectories of immune cell differentiation. Monocytes develop in the bone marrow and can disseminate into peripheral tissue. In tissue-enrichment analyses of monocyte subsets, a C-X-C chemokine receptor type 4 ($CXCR4^{high}$)–to-$CCR2^{high}$ transition was discovered to accompany monocyte egress from the bone marrow to the peripheral tissues where $CCR2^{high}$ monocytes mature further into an interleukin-1β–expressing ($IL1B^+$) state. Furthermore, systematic comparisons of immune cell progenitors revealed a high abundance of B cell progenitors across almost all prenatal organs, in contrast to the thymus-restricted distribution of T cell progenitors. This finding builds on the previous notion that hematopoiesis is limited to certain tissues such as the liver, bone marrow, and yolk sac and requires further research.

These cross-tissue atlases could also help reveal cell types that underlie disease pathogenesis. Although numerous disease-associated genes have been identified, the cellular context and tissue environment in which these genes function are often opaque. Eraslan *et al.* analyzed muscle diseases as examples of monogenic disorders in which mutations occur in myocyte-expressed genes. By relating these genes to cardiac, skeletal, and smooth muscle cells across a wide range of tissue types, the authors identified finer myocyte subsets that malfunction. For example, myocytes localized at the neuromuscular junction, but not those in other locations, are enriched with the expression of genes related to a type of neuromuscular transmission disease. Additionally, ligand-receptor analyses between disease-associated myocytes and other cell types across tissues highlighted interrupted cellular interactions as a mechanism of muscular disease pathogenesis. For instance, mutations in the receptor tyrosine protein kinase ERBB3 may disrupt ligand-receptor signaling between myocytes and Schwann cells (which produce the myelin sheath around peripheral neurons) in a type of lethal muscular dysfunction.

For complex diseases, loci that were identified in genome-wide association studies (GWASs) were enriched in specific cell types involved in the tissue of disease manifestation and, intriguingly, were also enriched in similar cell types residing in tissues unaffected by the disease. This is exemplified by the finding that coronary heart disease and heart rate loci were enriched in pericytes from breast, esophagus, and skeletal muscle in addition to the heart. These findings warrant further investigation of tissue microenvironments and how the overall cellular interactions combine to induce disease pathogenesis.

These pan-tissue human cell atlases form important reference datasets for understanding and predicting the side effects and safety issues of new medicines. Despite the increasingly targeted nature of new therapeutics, these agents are typically administered systemically, raising concerns about off-target effects (*10*). For example, $CD19^+$ neuronal cells in the brain are associated with brain toxicities of CD19-targeted chimeric antigen receptor (CAR)–T cell therapies (*11*). Ideally, the on-target, off-tissue side effects can be predicted and monitored, instead of being discovered in the clinic. Thus, a pan-tissue query of the single-cell expression of a therapeutic target would empower toxicity predictions before the first-in-human experiment.

Collectively, these pan-tissue studies bring us closer to building a comprehensive human single-cell atlas. Future analyses may include larger and diverse cohorts. It is also crucial to apply this approach in the setting of disease, such as cancer. Understanding both shared and tissue-specific features of tumorigenesis is key to the development of histology-agnostic and cancer type–specific therapeutics. For example, immunotherapies using programmed cell death protein 1 (PD1) antibodies have shown efficacy for multiple cancer types. Recently, pan-cancer single-cell analyses of tumor immune microenvironments have been reported for myeloid cells (*12*) and T cells (*13*) and should be broadened to include more comprehensive cell types and patient cohorts. Such initiatives will certainly spur a wave of technical advancements and biological insights. Ultimately, future understanding of human biology should encompass individual genes, cell types, and phenotypes in an integrated manner. ■

REFERENCES AND NOTES

1. K. J. Travaglini *et al.*, *Nature* **587**, 619 (2020).
2. L. W. Plasschaert *et al.*, *Nature* **560**, 377 (2018).
3. P. K. L. Murthy *et al.*, *Nature* **604**, 111 (2022).
4. E. A. Winkler *et al.*, *Science* **375**, eabi7377 (2022).
5. R. K. Perez *et al.*, *Science* **376**, eabf1970 (2022).
6. The Tabula Sapiens Consortium, *Science* **376**, eabl4896 (2022).
7. G. Eraslan *et al.*, *Science* **376**, eabl4290 (2022).
8. C. Domínguez Conde *et al.*, *Science* **376**, eabl5197 (2022).
9. C. Suo *et al.*, *Science* 10.1126/abo0510 (2022).
10. A. M. Luoma *et al.*, *Cell* **182**, 655 (2020).
11. K. R. Parker *et al.*, *Cell* **183**, 126 (2020).
12. S. Cheng *et al.*, *Cell* **184**, 792 (2021).
13. L. Zheng *et al.*, *Science* **374**, abe6474 (2021).

ACKNOWLEDGMENTS

The authors are supported by the Beijing Municipal Science and Technology Commission (Z201100005320014), the National Natural Science Foundation of China (81988101, 91942307, and 31991171), and the Ministry of Science and Technology of China (2020YFE0202200).

10.1126/science.abq2116

GRAPHIC: A. MASTIN/*SCIENCE*

MICROBIOLOGY

Recording bacterial responses to changes in the gut environment

A CRISPR-based tool reveals intestinal microbiota gene expression through time

By **Liron Zahavi**[1,2] and **Eran Segal**[1,2]

The bacterial population in the human gut is tightly linked with host health and has been implicated in a wide array of diseases, such as atherosclerosis, diabetes, and cancer (*1*, *2*). Existing treatments that are intended to shape the microbiota to improve health—prebiotics, probiotics, and fecal microbiota transplantations (FMTs)—have limited success in obtaining the desired microbial composition and in maintaining it over time (*3*, *4*). This is largely because bacteria in the gut are affected by other bacteria, human physiology, diet, medications, and more, and the relationships in this complex system are yet to be deciphered. Understanding the bacterial response and adaptation to these variables is essential for creating more effective microbiota-based interventions. On page 714 of this issue, Schmidt *et al.* (*5*) present a new tool that "documents" bacterial gene expression in mice and can clarify the bacterial response to perturbations in the intestinal environment.

Most studies that investigate the effect of different factors on the gut microbiome are based on DNA profiling of fecal samples. Such studies compare the DNA content of microbiomes and associate differences with environmental perturbations. This approach reveals changes in species composition or in bacterial genomics, reflecting the microbiome response on ecological and evolutionary time scales, respectively. However, the most immediate bacterial response happens at a physiological time scale, which is not reflected in DNA (*6*). One immediate means of adaptation to environmental changes is gene expression. Thus, RNA-based studies fill an important gap for understanding bacterial adaptation to host variables.

Schmidt *et al.* introduce a method for studying microbial transcription. They demonstrate this method by colonizing mice with an engineered *Escherichia coli* strain, inserting the Record-seq system, which leverages the recording nature of CRISPR-Cas (*7*). Within the Record-seq–engineered *E. coli* cells, transcripts are reverse transcribed to DNA and stored as spacers in a CRISPR array. As transcriptional records are stored as DNA sequences in the CRISPR array, bacterial gene expression history can then be sampled by DNA sequencing of host fecal samples (see the figure). They use this recording tool to investigate mechanisms of *E. coli* adaptation to varying host diets, cocolonization with another species, bacterial gene deletion, and host inflammation.

They uncovered, for example, multiple metabolic genes whose expression was increased when the mice were fed a less diverse diet. They hypothesized that these genes allow the bacteria to feed on host mucosal sugars and predicted that a bacterium lacking two of these genes will have a competitive disadvantage when the host is fed with a restricted diet, which they validate experimentally. This shows the potential of this method to highlight genes that allow bacteria to adapt to specific host conditions. These insights could be used to design an intervention that is based on bacteria that fit the host diet or to adjust host diet to allow certain bacteria to thrive. Similarly, they demonstrate the use of Record-seq to analyze the effect of introducing another species on *E. coli* gene expression and deduce that the bacteria shift their metabolism to exploit new niches created by the other species. Such experiments can be informative about bacteria-bacteria interactions in a mechanistic resolution that species co-occurrence measurements cannot.

The ability to look at microbial gene expression is not new. RNA-sequencing (RNA-seq)–based metatranscriptomics is a common tool with which to investigate the microbiome (*8*). There are, however, considerable advantages to Record-seq over RNA-seq. Although RNA-seq shows the immediate response of gut bacteria to its environment, it only reflects a snapshot of the microbiome at the moment of the sample. Because of the short lifetime of mRNA molecules, samples need to be taken at the exact moment of the response to investigate transient states. Record-seq, however, is distinctive in that it archives transcripts in the engineered *E. coli* DNA over time and thus describes a series of states. When Schmidt *et al.* changed the diets fed to mice, Record-seq identified signs of the previous diet after more than a week, whereas the signal was rapidly lost with RNA-seq analysis of fecal samples.

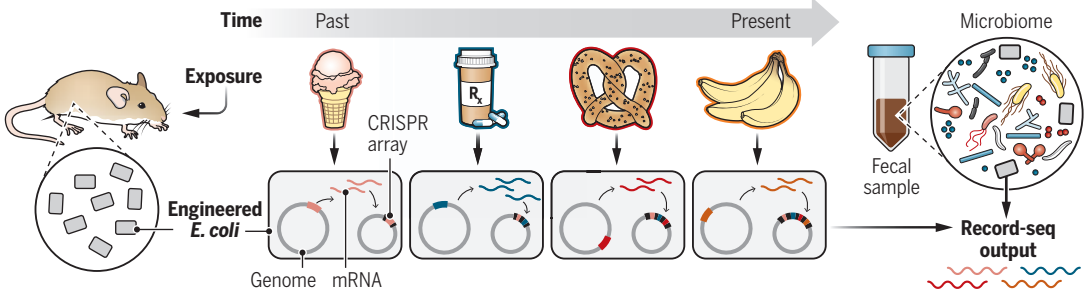
In a different experiment, Schmidt *et al.* demonstrate another advantage of their method: the ability to attribute transcripts to a specific CRISPR array and, therefore, bacterial strain. In metatranscriptomics, RNA is extracted from fecal samples and represents the wide array of species colonizing the gut. Attributing mRNA molecules to genomes depends on the specificity of the short mRNA sequence to a single genome. Schmidt *et al.* exemplify the importance of this limitation. They barcoded two CRISPR arrays and transformed the recording plasmids into *E. coli* cells of two strains. One was a wild-type strain, and the other had one gene deleted. They cocolonized the two strains in mice and studied the effect of the deletion

[1]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. [2]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. Email: eran.segal@weizmann.ac.il

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

Recording microbial responses in vivo

Schmidt *et al*. engineered *Escherichia coli* cells to record their gene expression in the mouse intestine. Within the engineered cells, the Record-seq system incorporates reverse-transcribed transcripts as new spacers in a barcoded CRISPR array on plasmids. Sequencing mice fecal samples revealed the bacterial response to ongoing and past exposures, such as diet changes and intestinal inflammation.



on gene expression of that strain compared with the wild type, showing the dynamics between the two similar strains in vivo. This analysis could not have been done with metatranscriptomics.

The archive quality of Record-seq could be leveraged for clinical applications. In diabetes mellitus, following the patient's blood glucose concentration is important for assessing disease progression and the effectiveness of the treatment. Although blood glucose concentration indicates whether the patient is in hyperglycemia at the moment of sampling, there is a measure—percentage of glycated hemoglobin (%HbA1c)—that reflects the glycemic state of the patient over the past 3 months. This test is invaluable for the diagnosis and management of diabetes, exemplifying the importance of a single test that captures longitudinal information. Other chronic diseases whose management depends on lifestyle adjustments can benefit from such a test. Schmidt *et al.* show that Record-seq–engineered sentinel cells can identify the existence and severity of intestinal inflammation in a colitis mouse model. Celiac patients, for example, could benefit from such a system of bacterial sentinel cells that monitors their adherence to dietary restrictions and the state of their disease. Furthermore, if successfully adapted to be used in human microbiota, this system can be used to design an assay for diet effects on inflammatory bowel disease (IBD) to develop a personalized nutritional plan to minimize intestinal inflammation in patients (*9*).

The method that Schmidt *et al.* introduce opens new avenues for studying the microbiota response to environmental factors and paves the way toward the development of more effective microbiota-based treatments. Two of the major advantages of their method—its ability to "document" responses in changing states and to associate outputs with specific strains within a complex community—are very much missing from other microbiota -omics research methods. Proteomics- and metabolomics-based research methods, which cover other important layers in microbiota dynamics and its interactions with the host, would benefit from similar improvements. ■

REFERENCES AND NOTES

1. S. V. Lynch, O. Pedersen, *N. Engl. J. Med.* **375**, 2369 (2016).
2. V. Gopalakrishnan, B. A. Helmink, C. N. Spencer, A. Reuben, J. A. Wargo, *Cancer Cell* **33**, 570 (2018).
3. J. Suez, N. Zmora, E. Segal, E. Elinav, *Nat. Med.* **25**, 716 (2019).
4. S. S. Li *et al.*, *Science* **352**, 586 (2016).
5. F. Schmidt *et al.*, *Science* **376**, eabm6038 (2022).
6. A. H. Yona, I. Frumkin, Y. Pilpel, *Cell* **163**, 549 (2015).
7. F. Schmidt, M. Y. Cherepkova, R. J. Platt, *Nature* **562**, 380 (2018).
8. Y. Zhang *et al.*, *Annu. Rev. Biomed. Data Sci.* **4**, 279 (2021).
9. C. B. Collins *et al.*, *Crohn's Colitis 360* **2**, otaa055 (2020).

10.1126/science.abq1455

MATERIALS SCIENCE

Beating natural proteins at filtering water

Artificial fluorous channels outperform aquaporins in water permeation

By **Yuexiao Shen**

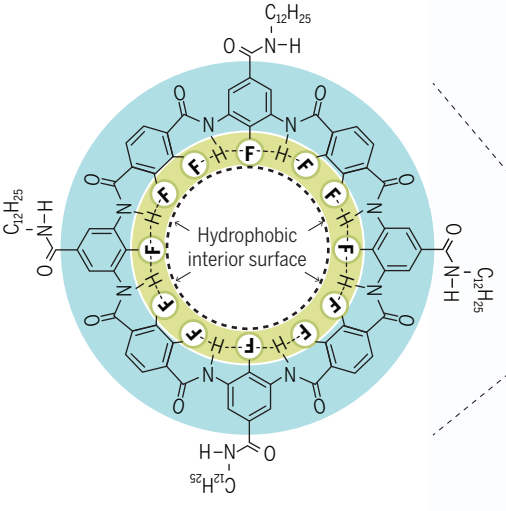
Humans have long drawn inspiration from nature to create tools that can mimic a specific function. In recent decades, researchers have taken this philosophy to the nanoscopic level, such as in the creation of synthetic structures for water purification inspired by biological systems. Aquaporins are a type of naturally occurring protein known as water channels, which have notable water permeability and selectivity (*1*). On page 738 of this issue, Itoh *et al.* (*2*) report the design of fluorous channels—structures made out of fluorocarbons—that can self-assemble from oligoamide nanorings and outperform biological water channels in permeability. Because of the superhydrophobic interior surface endowed by densely covered fluorine atoms, these nanochannels exhibit a water permeation flux that is two orders of magnitude greater than that of aquaporins.

Nature evolves ideal water-selective channels, including aquaporins, that are embedded in cell membranes to regulate osmotic balance and make critical biophysical contributions to living organisms. Aquaporins form an hourglass-shaped narrow channel with a ~0.3-nm aperture and a hydrophobic interior surface, resulting in notable water permeability and solute rejection (*1*). However, these energy-efficient structures have been difficult to mimic. The first artificial water channels could accommodate water molecules, but the mobility was slow and there was little rejection of other small molecules (*3*, *4*). Later synthetic structures maintained high water conductance (*5*–*8*) while also rejecting salts, including NaCl (*6*–*9*), and even protons (*7*, *8*).

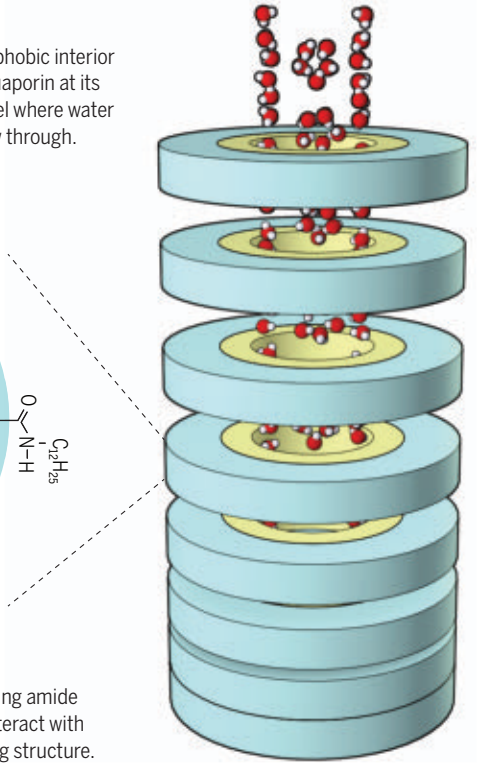
When water flow is confined to a subnanometer channel with hydrophobic interior surfaces, classical hydrodynamics is no longer applicable. The clustered arrangement of water molecules breaks

A frictionless water channel

Itoh *et al.* designed a nanochannel with a superhydrophobic interior surface that beats the naturally occurring protein aquaporin at its own game. The stacked nanorings form a nanochannel where water clusters are broken down into a single file as they flow through.



The ring is fabricated by condensing fluorine-containing amide monomers. The electronegative fluorine (F) atoms interact with the adjacent hydrogen atoms, which stabilizes the ring structure.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

down and leads to the formation of single-file water "wires" within channels. This single-file arrangement, in combination with an increase of frictionless "slip flow" at the channel walls, is believed to be the key factor that enables aquaporins and their mimics to achieve extremely fast water flow (*10*). Itoh *et al.* created a new class of synthetic channels that was inspired by the properties of Teflon, a well-known superhydrophobic fluorous material. They developed a series of fluorine-based macrocyclic channels whose water permeability exceeds that of aquaporins by more than 100-fold.

To create confined fluorinated surface chemistry at the nanoscale, Itoh *et al.* oligomerized fluorine-containing amide monomers through condensation reactions (see the figure). They report that because of their high electronegativity, fluorine atoms electrostatically interact with the hydrogen atoms from the adjacent polar amide groups and form hydrogen bonds. This makes the carbon-fluorine bonds point inward and the entire macrocyclic ring rigid, resulting in a series of fluorous oligoamide nanorings with pore diameters ranging from 0.9 to 1.95 nm. After these rings are stacked on top of each other within a lipid membrane, the nanochannel's interior surface is densely coated with fluorine atoms.

"They developed a series of fluorine-based macrocyclic channels whose water permeability exceeds that of aquaporins by more than 100-fold."

The superhydrophobic fluorinated surface enhances the breakdown of water clusters, as demonstrated by molecular dynamics simulations, thus substantially improving the water transport rate. Not surprisingly, these fluorous channels were experimentally shown to exhibit two to three orders of magnitude greater water permeability (5.5×10^{-10} cm^3 per second per channel) than both aquaporins and recent aquaporin mimics, including synthetic channels and carbon nanotubes. In addition to exceptional water permeability, these fluorous channels also offer excellent rejection of anions such as chloride because of the electrostatically negative surface. This means that potential future applications such as desalination and water purification are possible. A key feature of aquaporins—proton exclusion combined with high water transport—was not demonstrated by Itoh *et al.*, but this is less relevant to water-purification applications.

It is exciting to see that by exercising exquisite control of synthetic chemistry, it is possible to engineer synthetic channels at the nanoscale to achieve superfast water transport and go beyond the performance of aquaporins. However, before this technology can be used in real-world applications, there are a few challenges to overcome. Even if these newly developed channels have exceptional transport properties, the low synthesis yield (<0.2%) could be a hurdle for scalable applications. There is also the practical challenge of incorporating the channels into filtration materials, that is, whether they can be used as additives in the traditional membrane fabrication process to enhance the performance (*11*) or whether their use requires a complete rethink of the manufacturing process. Researchers need to not only reengineer materials at the microscopic scale to improve macroscopic performance but also find solutions by combining material innovation with scalable approaches (*12*). If the goal is to make membrane separation materials, future research will need to focus on implementing the concept of fluorous channels in the structure of scalable polymer membranes, for example, to figure out how to promote water transport within the polymer matrix and enhance membrane toughness such as chlorine resistance. Also, the field has converged on the tenet that selectivity is still the foremost aspect when considering synthetic channel design because critical separations in the subnanoscale range are enabled by increasing selectivity rather than permeability (*13*). The fluorous channels reported by Itoh *et al.* bring more diverse chemistry and possibilities for the design of synthetic channels and separation materials in the future. ■

Department of Civil, Environmental, and Construction Engineering, Texas Tech University, Lubbock, TX 79409, USA. Email: yuexiao.shen@ttu.edu

REFERENCES AND NOTES

1. P. Agre, *Angew. Chem. Int. Ed.* **43**, 4278 (2004).
2. Y. Itoh *et al.*, *Science* **376**, 738 (2022).
3. Z. Fei *et al.*, *Angew. Chem. Int. Ed.* **44**, 5720 (2005).
4. X. B. Hu *et al.*, *J. Am. Chem. Soc.* **134**, 8384 (2012).
5. R. H. Tunuguntla *et al.*, *Science* **357**, 792 (2017).
6. W. Song *et al.*, *Nat. Nanotechnol.* **15**, 73 (2020).
7. Z.-J. Yan *et al.*, *J. Am. Chem. Soc.* **142**, 15638 (2020).
8. A. Roy *et al.*, *Nat. Nanotechnol.* **16**, 911 (2021).
9. E. Licsandru *et al.*, *J. Am. Chem. Soc.* **138**, 5403 (2016).
10. W. Song, M. Kumar, *Curr. Opin. Chem. Eng.* **25**, 9 (2019).
11. M. Di Vincenzo *et al.*, *Nat. Nanotechnol.* **16**, 190 (2021).
12. E. Abaie *et al.*, *Front. Environ. Sci. Eng.* **15**, 124 (2021).
13. H. B. Park *et al.*, *Science* **356**, eaab0530 (2017).

10.1126/science.abo2953

QUANTUM PROCESSES

Anomalous fluid flow in quantum systems

New regimes of hydrodynamics are probed with synthetic quantum matter

By **Alan Morningstar** *and* **Waseem Bakr**

Recent progress in cooling, controlling, and measuring a small number of isolated atoms or ions has led to the creation of experimental platforms with which quantum phenomena can be realized and explored in fine detail. In quantum simulation experiments, idealized theoretical models are implemented on these platforms to study them experimentally in otherwise inaccessible regimes. For example, a set of many interacting quantum "spins" can be simulated in the experiment, and the collective fluid-like dynamics they exhibit can be measured. The theoretical framework describing this fluid flow, known as hydrodynamics, is said to "emerge" from the basic laws that govern the system. How exactly this happens, and what emergent dynamics are possible, are questions at the frontier of physics. On pages 720 and 716 of this issue, Joshi *et al.* (*1*) and Wei *et al.* (*2*), respectively, report on observations of unusual types of hydrodynamics in quantum simulations.

Despite a difference in the underlying experimental platforms, both experiments are simulations of one-dimensional (1D) systems of quantum spins, each with two states: up and down. Pairs of spins can exchange their states by interacting but cannot independently flip. As a result, the difference in the total number of up spins and down spins—the total magnetization—does not change under the dynamics.

However, as a collective effect, magnetization does flow within these systems. The most common such hydrodynamics is called diffusion, which is the type of flow

Department of Physics, Jadwin Hall, Princeton University, Princeton, NJ 08544, USA. Email: aorm@princeton.edu

that results from spin moving through the system with random short-range hops. In diffusion, a local excess of magnetization decays slowly with time according to a dynamical exponent, commonly denoted as the letter z, with a value of 2. By contrast, a system in which spin streams through uninterrupted is called "ballistic" and hosts a faster decay, with a z value of 1. Joshi *et al.* and Wei *et al.* present different experiments in which anomalous hydrodynamics and intermediate "superdiffusion" are observed, corresponding to values of z between 1 and 2 (see the figure).

Joshi *et al.* use a chain of trapped ions to simulate a 1D spin system with long-range spin exchange interactions. The parameter α, which characterizes the decay of the strength of the long-range interactions as a function of distance, is tunable in their experiment. Because of these interactions, one could expect the hydrodynamics of magnetization to be the same as in a system in which spin randomly hops around, similar to diffusion, but with a distribution of hop sizes that matches the long-range interactions. This behavior is known as a "Lévy flight" (*3*). In Lévy flights, when large hops are common enough, they become relevant and cause the flow of magnetization to become anomalous, with a z value that is <2 and changes as the parameter α is tuned (*4*).

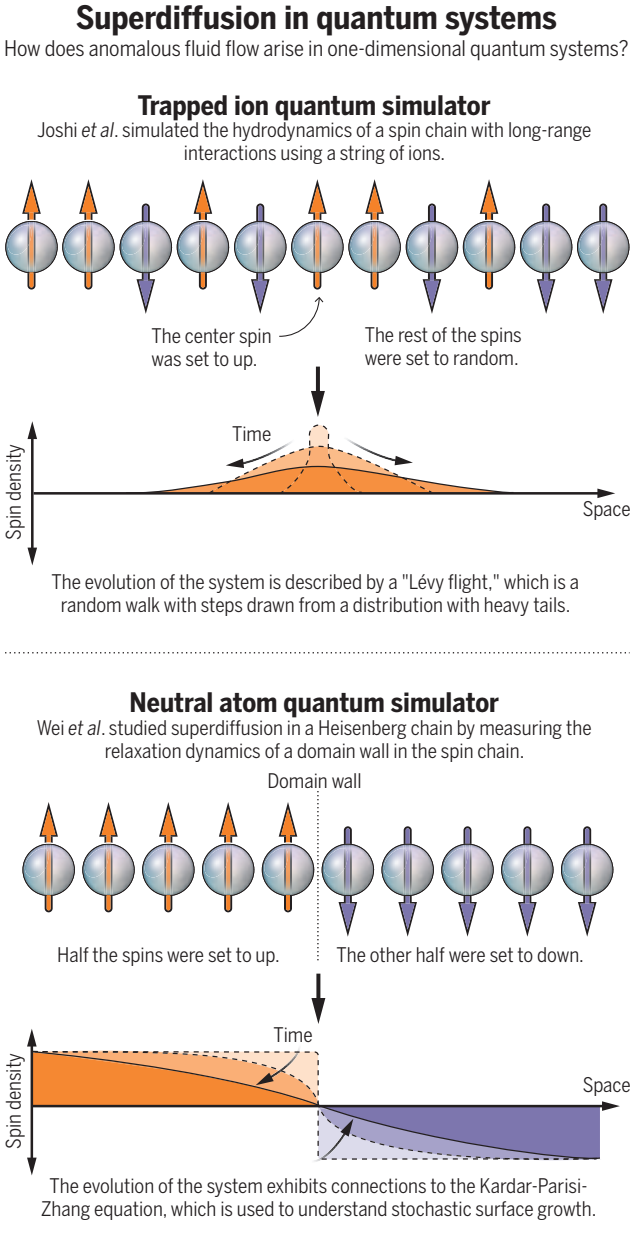
To test the Lévy flight theory, Joshi *et al.* prepare a set of initial states with all spins in the chain set randomly up or down, except for the spin at the center of the chain, which is initially set to up. Using this setup, they measure how the central excess magnetization spreads through the chain over time. This is done for several values of the interaction parameter α, and the dependence of the measured dynamical exponent z on α is consistent with the Lévy flight picture.

In a separate work, Wei *et al.* use ultracold atoms to simulate a so-called "Heisenberg" spin chain with short-range interactions. The initial states they prepared were "domain walls," in which one half of the chain is spin up and the other half is spin down. As spin is exchanged between the two halves and the system approaches equilibrium, the authors measure the net magnetization transferred across the initial boundary, which allows them to characterize the flow of magnetization in the system.

The Heisenberg spin chain has a certain rotational symmetry and a special property known as integrability. These two ingredients are predicted to result in a specific type of superdiffusion with the dynamical exponent z of 3/2. Wei *et al.* demonstrate this superdiffusion in their system and probe it further by destroying it in two distinct ways. First, they tune the total magnetization away from zero, breaking the rotational symmetry, which causes the system to become ballistic: z goes to 1. Then, they leave the symmetry intact but break the integrability, which leads to diffusion: z goes to 2. In this way, the authors provide experimental evidence that it is the combination of integrability and symmetry that gives rise to superdiffusion in the Heisenberg spin chain (*5*). Exploration of these dynamics is motivated by an unexpected connection to a class of systems called the Kardar-Parisi-Zhang (KPZ) universality class, which describes stochastic surface growth dynamics but has no obvious relation to quantum dynamics (*6*–*11*).

Both Joshi *et al.* and Wei *et al.* show that quantum simulation experiments are a valuable approach to exploring the dynamics of quantum systems. Each set of authors observe and characterize an anomalous emergent hydrodynamics and are able to tune its properties by varying the underlying experimental parameters. Although both experiments result in superdiffusion, the underlying mechanisms are shown to be very different. In the case of Joshi *et al.*, the mechanism seems to be well understood, and the experiment is a crisp realization of emergent Lévy flight–like dynamics in a quantum system, whereas in Wei *et al.*, there is still some mystery surrounding the connection to KPZ.

Although hydrodynamics is a nonquantum framework, the underlying properties that allow these experiments to be done are quantum mechanical. It will be interesting to see what other types of emergent dynamics will arise under the "quantum conditions" found in modern quantum simulation experiments. As these simulators continue to push past the capabilities of first-principles theoretical methods, they are sure to move the study of quantum dynamics into an exciting data-rich regime. ■


Superdiffusion in quantum systems
How does anomalous fluid flow arise in one-dimensional quantum systems?
Trapped ion quantum simulator
Joshi et al. simulated the hydrodynamics of a spin chain with long-range interactions using a string of ions.
The center spin was set to up.
The rest of the spins were set to random.
Time
Spin density
Space
The evolution of the system is described by a "Lévy flight," which is a random walk with steps drawn from a distribution with heavy tails.
Neutral atom quantum simulator
Wei et al. studied superdiffusion in a Heisenberg chain by measuring the relaxation dynamics of a domain wall in the spin chain.
Domain wall
Half the spins were set to up.
The other half were set to down.
Time
Spin density
Space
The evolution of the system exhibits connections to the Kardar-Parisi-Zhang equation, which is used to understand stochastic surface growth.

REFERENCES AND NOTES

1. M. K. Joshi *et al.*, *Science* **376**, 720 (2022).
2. D. Wei *et al.*, *Science* **376**, 716 (2022).
3. V. Zaburdaev, S. Denisov, J. Klafter, *Rev. Mod. Phys.* **87**, 483 (2015).
4. A. Schuckert, I. Lovas, M. Knap, *Phys. Rev. B* **101**, 020416 (2020).
5. M. Dupont, J. E. Moore, *Phys. Rev. B* **101**, 121106 (2020).
6. M. Znidarič, *Phys. Rev. Lett.* **106**, 220601 (2011).
7. M. Ljubotina, M. Žnidarič, T. Prosen, *Nat. Commun.* **8**, 16117 (2017).
8. M. Ljubotina, M. Žnidarič, T. Prosen, *Phys. Rev. Lett.* **122**, 210602 (2019).
9. M. Kardar, G. Parisi, Y.-C. Zhang, *Phys. Rev. Lett.* **56**, 889 (1986).
10. V. B. Bulchandani, S. Gopalakrishnan, E. Ilievski, *J. Stat. Mech.* **2021**, 084001 (2021).
11. A. Scheie *et al.*, *Nat. Phys.* **17**, 726 (2021).

10.1126/science.abn6376

GRAPHIC: V. ALTOUNIAN AND N. CARY/*SCIENCE*

POLICY FORUM

CLIMATE POLICY

Investor-state disputes threaten the global green energy transition

Global action on climate change could generate upward of $340 billion in legal claims from oil and gas investors

By **Kyla Tienhaara[1,2], Rachel Thrasher[2], B. Alexander Simmons[2], Kevin P. Gallagher[2]**

To limit global warming to below 1.5°C, governments will have to simultaneously curb demand for fossil fuels and limit supply (*1*). However, efforts to limit supply will affect asset holders, particularly in the upstream (exploration and production) and midstream (transportation and storage) portions of the supply chain. Demands for compensation will ensue, and when the companies involved are foreign, legal claims may be brought to international arbitration in a process known as investor-state dispute settlement (ISDS) (see the first figure). The Intergovernmental Panel on Climate Change (IPCC) has recently acknowledged that ISDS cases could lead to states refraining from, or delaying, measures to phase out fossil fuels (*2*). Here, we illustrate legal and financial risks associated with limiting oil and gas production and argue that governments should take steps to prevent fossil fuel investors from accessing ISDS.

The International Energy Agency (IEA) Net-Zero Emissions Scenario (NZE) by 2050 has no new investments in fossil fuel supply after 2021 (*3*). In a 1.5°C carbon budget, nearly 60% of oil and fossil methane gas reserves and 90% of coal reserves cannot be extracted (*4*), suggesting that even some existing operations are unviable. Although many government plans for fossil fuel production are inconsistent with their commitments under the Paris Agreement (*5*), the launch of the Beyond Oil and Gas Alliance (BOGA) in 2021, by 11 national and subnational governments pledging Paris-aligned dates for ending oil and gas production, suggests that the tide may be turning. At the same time, a UK firm, Rockhopper, is using ISDS to seek an undisclosed amount over Italy's denial of a coastal drilling permit, and a Canadian firm, TC Energy, is claiming more than $15 billion in compensation for US President Biden's cancellation of the Keystone XL Pipeline (table S1).

[1]School of Environmental Studies and Department of Global Development Studies, Queen's University, Kingston, ON, Canada. [2]Global Development Policy Center, Boston University, Boston, MA, USA. Email: kyla.tienhaara@queensu.ca

PROTECTIONS FOR FOREIGN INVESTORS

Thousands of bilateral and plurilateral investment treaties have been signed by states over the last 50 years (*6*). Investment treaties act as a form of political risk insurance and were presented to governments in the Global South as a means to increase foreign investment flows. Under these treaties, foreign investors are protected from certain forms of state conduct but have no responsibilities themselves. Investors are empowered to bring claims that a state has breached a treaty directly to an ad hoc ISDS tribunal composed of three members—one chosen by the investor, one by the state, and one that is mutually agreed upon or appointed by an arbitration body such as the International Centre for the Settlement of Investment Disputes (ICSID).

Foreign investors have relied on various investment treaty rules to claim that policies to protect the environment, public health, or the economy are discriminatory, undermine their "legitimate expectations," or impose a burden that is not justified or proportional to the state's policy objectives (*7*) (table S2).

If it is determined that a state has breached an investment treaty, the tribunal will typically require the government to pay the investor compensation. Increasingly, there is a tendency for tribunals to use projections of an investment's expected future income across its entire life cycle as the basis of compensation, using discounted cash flow as the method of calculation (*8*). The result has been hugely inflated monetary awards in the hundreds of millions and even billions of dollars. In some cases, the amount of compensation awarded has been described as "crippling" because it is incommensurate to the state's capacity to pay (*9*). Governments don't always lose, but the rules are sufficiently vague to make it difficult to predict outcomes. This, combined with the risk of a very large award, can create a chilling effect. Critically, despite a growing body of research (*10*), there is a lack of understanding on the part of many states that the treaties do not live up to their purported benefit of facilitating investment.

ISDS AND THE FOSSIL FUEL SECTOR

Law firms are advising investors that are "most affected by States' climate change obligations" (i.e., those in the fossil fuel sector) to (re)structure their investments to ensure access to ISDS (*11*). To date, there have been at least 231 ISDS cases related to fossil fuel investments, which is almost 20% of the total known number of cases across all sectors (*12*). The vast majority (92%) of fossil fuel cases are related to oil and gas investments (*12*). Fossil fuel investors have been successful in

The stages of an investor-state dispute settlement

Foreign investors may have a legal claim for compensation if countries cancel treaty-protected fossil fuel projects where investors have already obtained at least an exploration permit. If the lawsuit is successful, countries face substantial financial losses and may pull out of climate commitments to avoid additional ISDS claims from other investors. Colors of each stage correspond to the primary actor(s) involved: the investor (red), state (blue), tribunal (gray), or multiple actors (purple).

Investment treaties and concessions
States sign investment treaties to encourage foreign investment and allocate concessions to fossil fuel investors years in advance of the investment's operation.

Action from the state
The state decides to phase out future or existing fossil fuel–based projects to align with their new climate commitments.

Notice of arbitration
Foreign investor of an affected project notifies the state of a potential breach of its commitments under an international investment agreement.

Selection of arbitrators
Three lawyers are appointed by the state and foreign investor as arbitrators in the investor-state dispute settlement (ISDS).

Tribunal proceedings
Arbitrators examine if the state undermined the investor's "legitimate expectations" or the value of their investment and estimate losses, stranded assets, or future profits.

Judgment rendered
Arbitrators determine if the state violated its previous commitments and owes compensation to the injured foreign investor.

Troubling feedbacks
Expensive awards or settlements could lead to states pulling out of climate commitments and rolling back regulations out of fear of more ISDS cases.

GRAPHIC: K. FRANKLIN/*SCIENCE* BASED ON K. TIENHAARA *ET AL.*

72% of cases that were decided on the merits where the final award was disclosed (*12*). Fossil fuel cases have a high average amount of compensation awarded ($600 million) (*12*), and asset valuation is complicated by the volatility of oil prices, which are highly sensitive to exogenous shocks such as the pandemic and the war in Ukraine.

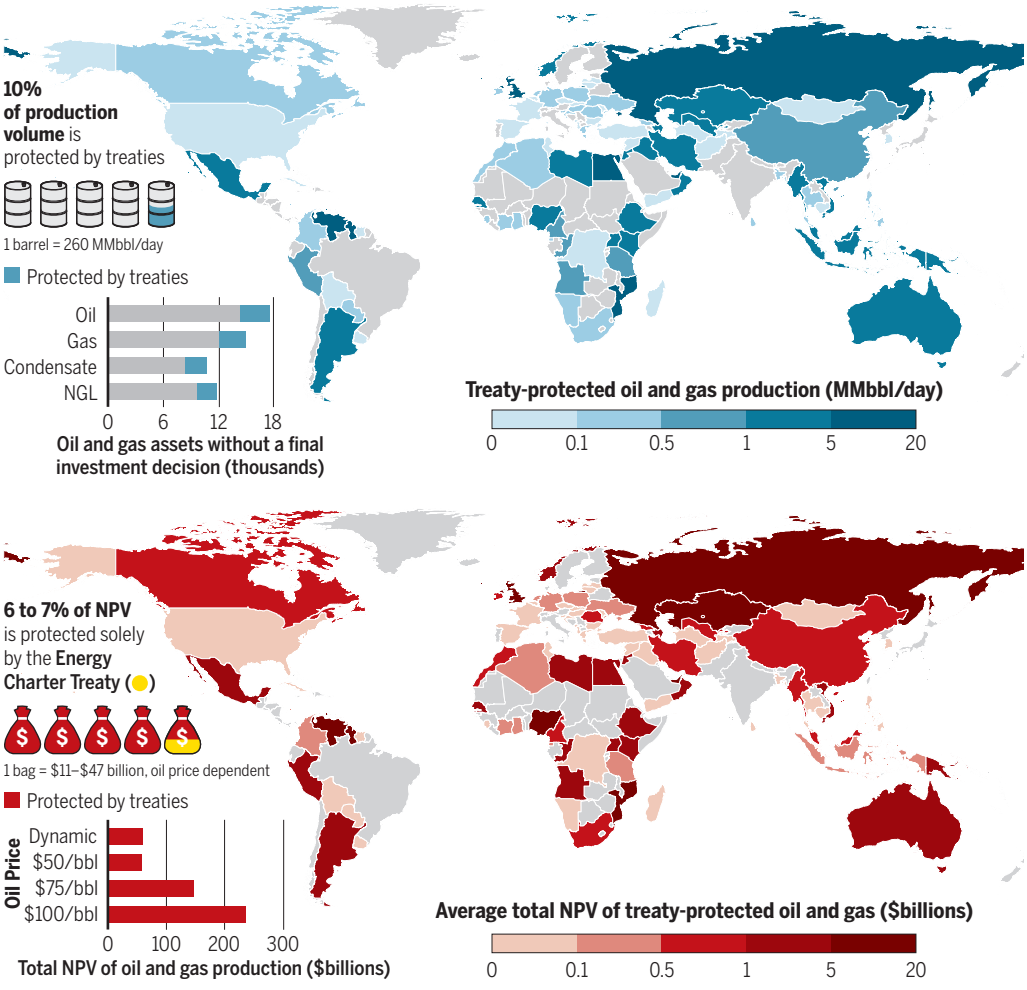
Although only a small number of ISDS cases to date concern climate actions (table S1), this will likely change if governments begin to adopt more stringent policies, particularly if those policies directly affect fossil fuel investors (table S3). States can argue that investors should have anticipated that they would be affected by climate policies, given long-standing scientific consensus that fossil fuels are the primary source of greenhouse gas emissions and international agreements on climate change mitigation dating back 30 years. Nevertheless, it remains unclear how tribunals will weigh generalized commitments by states to address climate change against specific commitments to further oil and gas production (e.g., through issuance of leases).

Countries with a large share of fossil fuel assets under the protection of foreign investment treaties are faced with substantial legal and financial risks if they phase out fossil fuel production. For example, if states choose to follow the NZE and cancel all fossil fuel projects that do not have a final investment decision as of 31 December 2021 (*3*), foreign investors that have already obtained at least an exploration permit by this date would have grounds for an ISDS claim (as in *Rockhopper* v. *Italy* and *Lone Pine* v. *Canada*; table S1). States could take more aggressive action and additionally cancel projects that have a final investment decision but are not yet producing (i.e., under development), which could give rise to more ISDS claims.

To analyze risks in these scenarios, considering both the existing caseload of fossil fuel industry ISDS claims and the availability of comprehensive project-level data, we chose to focus exclusively on upstream oil and gas. We utilized Rystad Energy's UCube database, which contains a list of countries' crude oil, gas, condensate, and natural gas liquid (NGL) assets and their attributes, including the production status, real or anticipated approval year, country headquarters of foreign investors, production volume per investor, and estimates of the net present value (NPV) per investor. We merged this dataset with information on 1828 investment treaties that provide access to ISDS (*6*). To illustrate the potential range of financial compensation for treaty-protected projects under uncertainty of future oil and gas prices, we used four NPV estimates: one using Rystad Energy's temporally dynamic NPV prediction, and three assuming a constant price of industry-standard Brent oil (and gas equivalent) of $50, $75, and $100 per barrel (see supplementary materials for detailed methods).

Treaty-protected fossil fuel projects could delay climate action

(Top) Extent of the world's oil and gas projects without a final investment decision that are protected by foreign investment treaties, and their estimated production volume in millions of barrels per day (MMbbl/day). (Bottom) Estimated total net present value (NPV) of all treaty-protected oil and gas projects without a final investment decision across four oil price scenarios, including the average total NPV per country across all scenarios, and the share of the valuation protected exclusively by the Energy Charter Treaty.



BILLIONS IN COMPENSATION

UCube lists 55,206 upstream oil and gas projects in 159 countries that are predicted to have a final investment decision made during 2022–2050 but for which at least an exploration permit had already been obtained by investors. If countries cancel these projects in line with the NZE, ISDS claims are possible for at least 10,506 (19%) of these projects across 97 countries, which are currently protected by 334 different investment treaties. This represents 18 to 21% of all crude oil, gas, condensate, and NGL assets and 8 to 11% of the total estimated production of all oil and gas projects awaiting a final investment decision (see the second figure). The total NPV of these protected projects is expected to reach $60 billion to $234 billion, depending on future oil prices. Valuations for crude oil production are the greatest ($20–114 billion), followed by gas ($20–90 billion), condensate ($7–21 billion), and NGL ($2–8 billion) (see the second figure). This translates to tremendous potential financial losses for countries across income levels if ISDS claims are successful. The five countries with the greatest potential losses

GRAPHIC: K. FRANKLIN/*SCIENCE* BASED ON K. TIENHAARA *ET AL.*

from ISDS are Mozambique ($7–31 billion), Guyana ($5–21 billion), Venezuela ($3–21 billion), Russia ($2–16 billion), and the United Kingdom ($3–14 billion).

If countries decide to also cancel oil and gas projects that are currently under development, this could introduce substantial additional financial losses from ISDS claims. Of the 1941 oil and gas projects under development in the last decade across 62 countries, 225 are protected under 81 foreign investment treaties. Although this represents a smaller share of all oil and gas projects (12%) and production volume (9%) under development, potential valuations remain high. The total estimated NPV ranges from an additional $32 billion to $106 billion, with gas ($12–52 billion) and crude oil ($14–42 billion) constituting the largest share. In this scenario, 33 countries would be vulnerable to additional ISDS claims, most notably Russia ($5–23 billion), Mozambique ($5–19 billion), Kazakhstan ($6–18 billion), Guyana ($2–6 billion), and Indonesia ($3–4 billion).

The Energy Charter Treaty (ECT) is the greatest contributor to these potential ISDS claims. The ECT, the only investment treaty with an exclusive focus on energy, has been ratified by 50 countries, mostly in Europe (table S4). The ECT applies to 19% of all treaty-protected oil and gas projects without a final investment decision, and 13% of those projects are only covered by the ECT. This represents 7% of the production volume of all treaty-protected projects and 6 to 7% ($3–16 billion) of the total NPV of those projects. For oil and gas projects that are under development, 6% of all treaty-protected projects are only covered by the ECT, which constitutes 4 to 6% ($2–4 billion) of the total valuation.

AN ABOLITIONIST APPROACH

International treaties that provide protection to investors and access to ISDS increase the power of the fossil fuel industry to resist the implementation of policy measures that are crucial for the global green energy transition. When investors launch claims for compensation under ISDS, or threaten to do so, this can result in the delay of policy action, extension of deadlines for moratoria, or exemptions and payment of far greater compensation to leaseholders than would be available under domestic remedies. For example, fears of ISDS prevented New Zealand from becoming a full member of BOGA and committing to a Paris-aligned phase-out of oil and gas (*13*). Ultimately, countries must decide if the financial risks of immediate climate action outweigh the societal risks from delayed climate action. At the very least, governments should immediately halt oil and gas exploration. This would directly align with global climate targets, and set an immediate cap on the extent of ISDS compensation claims countries might face in the future.

More direct actions will be necessary to reduce existing ISDS threats without limiting national climate ambitions. The Organisation for Economic Co-operation and Development is conducting work on how this can be achieved (*14*). Governments should take measures to prevent existing leaseholders from accessing ISDS, such as terminating all bilateral investment treaties (*10*). Several countries, including South Africa, have done so, without any resulting reductions in foreign investment. States can also negotiate the removal of ISDS clauses from trade agreements, as the United States did with Canada in the US-Mexico-Canada Agreement. Another option is for states to withdraw consent to ISDS in cases involving fossil fuel investments, emulating the approach taken by Singapore and others to remove the threat of ISDS claims from the tobacco industry.

Our proposals are more efficient and effective than trying to fix the investment treaty system through procedural reforms or the introduction of exceptions clauses for climate policy. Reformist approaches would be time-consuming and likely ineffectual, based on the experience of previous efforts. Nevertheless, there are two major obstacles to our abolitionist approach. The first is "sunset clauses" that allow for continuation of treaties for 10 to 20 years for investments commenced prior to termination. Sunset clauses can be nullified through mutual agreement of the parties. The second obstacle—resistance to treaty termination or modification from states with powerful fossil fuel lobbies—could prove more difficult to overcome.

These issues are playing out in discussions around the "modernization" of the ECT. Given that the ECT protects far more oil and gas production than any other treaty, it should be prioritized for termination. This could reduce the global price tag of oil and gas project cancellations by $5 billion to $20 billion, or more. Yet although several European members are considering termination or coordinated withdrawal, 32 additional countries have begun the process to accede to the treaty (table S4). Expanding the ECT to these countries would increase the number and production volume of treaty-protected oil and gas projects without a final investment decision by 6% and 25%, respectively, and the total NPV of these new protected projects could increase the global ISDS price tag by 19 to 24% ($12–45 billion). Terminating the ECT would thus substantially reduce existing financial risks and eliminate additional future risks from expansion.

Although wealthy nations, such as the United Kingdom, Denmark, and Ireland, would benefit most from the avoided financial losses that would come with ECT termination, the threat of ISDS claims through other treaties for low- and middle-income countries would remain a major concern, particularly for countries vulnerable to severe climate change impacts and struggling to escape the debt crisis. The high end of our liability estimate ($340 billion) is more than the total level of public climate finance globally in 2020 ($321 billion) (*15*). Potential litigation risks are even greater if coal mining and midstream fossil fuel infrastructure are considered. We cannot afford to divert such a substantial amount of public finance from essential mitigation and adaptation efforts to compensate investors in fossil fuels. ■

REFERENCES AND NOTES

1. F. Green, R. Denniss, *Clim. Change* **150**, 73 (2018).
2. IPCC, *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, P. R. Shukla *et al*., Eds. (Cambridge Univ. Press, 2022).
3. IEA, *Net Zero by 2050: A Roadmap for the Global Energy Sector* (IEA, 2021).
4. D. Welsby, J. Price, S. Pye, P. Ekins, *Nature* **597**, 230 (2021).
5. Stockholm Environment Institute (SEI), International Institute for Sustainable Development (IISD), Overseas Development Institute (ODI), E3G, and UN Environment Programme (UNEP). "The Production Gap Report 2021"; http://productiongap.org/2021report.
6. United Nations Conference on Trade and Development (UNCTAD), "International Investment Agreements Navigator" (UNCTAD 2021); https://investmentpolicy.unctad.org/international-investment-agreements.
7. R. D. Thrasher, *Constraining Development: The Shrinking of Policy Space in the International Trade Regime* (Anthem, 2021), chap. 4, 7.
8. J. Bonnitcha, S. Brewin, "Compensation Under Investment Treaties" (IISD, 2020); https://www.iisd.org/system/files/publications/compensation-treaties-best-practicies-en.pdf.
9. M. Paparinskis, *ICSID Rev. Foreign Invest. Law J.* 10.1093/icsidreview/siab029 (2021).
10. C. Moehlecke, R. L. Wellhausen, *Annu. Rev. Polit. Sci.* **25**, 23.1 (2022).
11. Jones Day, "Climate Change and Investor-State Dispute Settlement" (Jones Day, February 2022); https://www.jonesday.com/en/insights/2022/02/climate-change-and-investorstate-dispute-settlement.
12. L. Di Salvatore, "Investor–State Disputes in the Fossil Fuel Industry" (IISD, 2021); https://www.iisd.org/system/files/2022-01/investor–state-disputes-fossil-fuel-industry.pdf.
13. E. Meager, "Cop26 targets pushed back under threat of being sued" (Capital Monitor, 14 January 2022); https://capitalmonitor.ai/institution/government/cop26-ambitions-at-risk-from-energy-charter-treaty-lawsuits/.
14. OECD, "The Future of Investment Treatiues"; https://www.oecd.org/investment/investment-policy/investment-treaties.htm.
15. B. Buchner *et al*., "Global Landscape of Climate Finance 2021" (Climate Policy Initiative, 2021); https://www.climatepolicyinitiative.org/publication/global-landscape-of-climate-finance-2021/.

ACKNOWLEDGMENTS

We acknowledge the Anishinaabe, Haudenosaunee, and Massachusett peoples, traditional owners of the lands on which we live and work, and pay our respects to their elders past, present, and emerging. We thank G. Muttitt, R. Ray, and anonymous reviewers for their feedback. This research was supported by the Open Societies Foundation, Rockefeller Brothers Fund, and the Canada Research Chairs Program.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo4637

Published online 5 May 2022
10.1126/science.abo4637

BOOKS *et al.*

Parker Roos, who has fragile X syndrome, dyes eggs with his family in 2012.

PHOTO: JIM YOUNG/REUTERS

HEALTH AND MEDICINE

Rethinking genetic disease

Carriers of the fragile X premutation can experience medical conditions of their own

By **Bill Sullivan**

It is widely assumed that being a carrier of a genetic disease means that only one's children are at risk of developing a medical condition. But there is at least one striking exception to this rule. The genetic aberration causing fragile X syndrome, which is marked by severe cognitive and behavioral problems, including autism, accumulates over time across generations. Seemingly unrelated medical issues have plagued parents and grandparents of individuals with fragile X syndrome. In *The Carriers*, psychiatrist Anne Skomorowsky tells their story and how doctors pieced together this complex hereditary puzzle.

The Carriers
Anne Skomorowsky
Columbia University Press, 2022. 280 pp.

Through absorbing case reports, Skomorowsky lifts the curtain on the bewildering array of maladies that fragile X carriers may endure. She was moved to learn more after seeing Jessie, a 22-year-old patient who was struggling with premature menopause and worried about her ailing father, who was developing tremors and having trouble walking. Surprisingly, a genetic test revealed that the pair's disparate medical problems were caused by a mishap at the same gene: fragile X mental retardation 1 (*FMR1*).

The *FMR1* gene makes a protein (FMRP) necessary for the proper development and function of neurons in the brain. In most patients with fragile X syndrome, the *FMR1* gene has been shut down, which means FMRP is no longer made.

Like all genes, *FMR1* is preceded by a stretch of DNA bases that regulate its activity. The DNA sequence upstream of *FMR1* contains repeats of three nucleotides (cytosine and two guanosines, or CGG). Most people have 30 to 40 CGG repeats preceding *FMR1*. Those harboring 55 to 200 CGG repeats are considered premutation carriers and are at risk of having children whose *FMR1* gene is epigenetically silenced by DNA methylation.

The Carriers focuses on the little-told story of those with the premutation, like Jessie and her father, who suffer from fragile X–associated primary ovarian insufficiency and tremor/ataxia syndrome, respectively. Skomorowsky deftly interweaves science and narrative, explaining how the variations in CGG repeats can lead to distinct, adult-onset disorders in different sexes and ages or produce "a touch of fragile X syndrome" in some carriers. She rightly points out that the difficulties suffered by some carriers, especially those who experience diminished executive function, make taking care of a fragile X child even more challenging.

The heroic efforts of astute researchers such as Randi Hagerman at the Medical Investigations for Neurodevelopmental Disorders (MIND) Institute at the University of California, Davis, who have sought to understand these conditions, read like a fast-paced detective story and have ramifications beyond medicine. Problems with behavior, aggression, and self-control, we are reminded, can stem from factors beyond an individual's control. Skomorowsky observes: "It wasn't just one damn thing after another. It was genes."

The chapter describing the so-called village of fools is among the more fascinating instances of the drastic impact that fragile X and its carriers can have on a community. The disproportionate amount of strange behavior and illness among the residents of Ricaurte, Colombia, is legendary, stretching back to its foundation in the 1800s. The medical mystery underlying this phenomenon was recently solved by physician Wilmar Saldarriaga Gil, whose keen sense of observation was second only to his compassionate curiosity.

Some may be disappointed that Skomorowsky does not evaluate potential future treatments for genetic diseases like fragile X and its premutations, although gene therapy has been covered extensively elsewhere. More mechanistic details into the function of FMRP and a description of the attempts underway to compensate for its deficiency with experimental therapeutics would have also been welcomed.

Two key takeaways emerge from Skomorowsky's enlightening account. The first is that fragile X and its associated premutations illustrate the important role that genomic analyses can play in understanding cognitive abilities and behavior, highlighting, in particular, that variations in intergenic regions—often dismissed as "junk DNA"—should not be ignored.

Second, *The Carriers* underscores the value of being patient with patients—attentive listening to their symptoms and family history can reveal clues needed to solve clinical mysteries. In this sense, patients and doctors work like two sides of the same coin: "When doctors listen closely to their patients, with enough care, they are not just being kind. They are part of a process of discovery." ■

10.1126/science.abn7477

The reviewer is the author of *Pleased to Meet Me: Genes, Germs, and the Curious Forces That Make Us Who We Are* (National Geographic Books, 2019). Email: wjsulliv@iu.edu

NEURODEVELOPMENT

The becoming of the human brain

A neuroscientist traces the development of the body's most complex organ

By **R. Douglas Fields**

Nothing in biology is more miraculous than the transformation of a fertilized egg into a complete organism. Contained within this process are all of genetics, the core of cell biology, and the imprints of evolution. The human brain in particular—formed from a teeming ball of embryonic cells—transcends anatomy to generate identity, personality, and comprehension, which make each new life different from all others. How the nervous system develops is the subject of neuroscientist William A. Harris's new book, *Zero to Birth: How the Human Brain Is Built*.

Using an expository style, interspersed with first-person accounts of his own research, Harris provides an authoritative synopsis of developmental neuroscience. The scope encompasses topics ranging from basic embryology to the influences of postnatal experience on brain development.

For developmental neuroscientists like myself, whose research careers stretch back to the 1980s, this accounting will serve as a delightful revisitation of the excitement surrounding noteworthy discoveries as they were made and the people who made them. For students and others new to the field, the book will provide a comprehensive compendium of the most important aspects of developmental neuroscience. Those new to the subject may find the proleptic beginning of the book a bit disorienting, but subsequent chapters follow a clear and orderly structure.

The book begins with the question of cell fate, opening with the classic experiments of Wilhelm Roux, Hans Driesch, and Hans Spemann, who worked to determine whether the twin cells formed by the first cleavage of a zygote would, if separated, produce hemiembryos or two complete twins in frogs, sea urchins, and salamanders, respectively. The answer—twins—formed the basis of our understanding of totipotency. Cells of this stage are capable of making it all. But this raises a grander question: What events lead a certain population of embryonic cells down the pathway to become the cells in our brains?

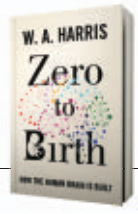
The narrative then jumps to genetic inheritance and single-cell organisms such as the paramecium to build an argument that "many properties that are characteristic of the brain were already encoded in the DNA of our single-cell ancestors." A brief tour of comparative neuroanatomy—from jellyfish to humans—comes next. Then the story detours to consider the fate of a single frog egg after fertilization. The remaining portion of the first chapter proceeds through the embryological milestones of forming a blastula, gastrulation, and the rise of the neural plate, which furrows and folds into the neural tube—the forerunner of the brain and spinal cord. Subsequent chapters discuss how the general body plan is formed, the proliferation and differentiation of neurons, and the broader topics of neural plasticity and human brain evolution.

Countless neural connections are refined as a baby's brain develops.

PHOTO: WAVEBREAKMEDIA/ISTOCK.COM

The reviewer is at the Section on Nervous System Development and Plasticity, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD 20892, USA. Email: fieldsd@mail.nih.gov

Zero to Birth: How the Human Brain Is Built
William A. Harris
Princeton University Press, 2022. 272 pp.



At this point, Harris confronts a seemingly unfathomable question: How do the roughly 100 billion neurons in the human brain get wired up properly? The specific labeling of each connection is mathematically impossible. Instead, the nervous system wires itself up by refining circuitry on the basis of performance. Countless connections in a baby's brain are winnowed away, its axons and dendrites pruned, and nubile neurons and nascent synapses that are less competitive in performance are eventually eliminated.

The final chapter considers the questions of what makes the human brain so different from all others and what makes your brain so different from every other human's. In large measure, the answer is found at the intersection of biology and society. Our brains continue to develop after birth, guided by the experiences we have in the environment we are born into.

The book covers a broad scope, but readers must look elsewhere for up-to-date information on the most abundant and diverse type of cells in the brain: glia. These cells have a profound influence on nearly every aspect of nervous system development, but they are referenced here only superficially. There is no mention of Schwann cells, for example, or of microglia. These and other glia help to guide neuronal development and prune synapses. In considering what makes the human brain different, Harris's omission of astrocytes, which in the human cerebral cortex are unlike those in any other animal, is another unfortunate oversight. On the subject of neuronal development, however, *Zero to Birth* is accessible, authoritative, and comprehensive. ■

10.1126/science.abp8642

Species in Himalayan headwater streams would be threatened by India's proposed small hydropower projects.

Edited by **Jennifer Sills**

India must protect Himalayan headwaters

At the 26th UN Climate Change Conference of the Parties (COP26), India pledged to convert to 50% renewable energy use by 2030 (*1*). The country plans to achieve this goal by building 5747 new small hydropower projects, more than 42% of which are planned in Indian Himalayan states (*2*). The 88 existing large dams–functional or under construction–in India's Himalayan region have already exacerbated habitat fragmentation and put local species and ecosystems at risk (*3*, *4*). Because the cumulative impacts of hundreds of small hydropower projects will likely lead to similar environmental damage, their construction must be reconsidered.

The Indian Himalaya's extraordinary vertebrate diversity (*5*) faces rapid development and associated habitat loss; 17.3% of the region's fish and amphibians are already threatened (*5*). Headwater aquatic species can only tolerate a narrow range of cold temperatures, making them particularly vulnerable to climate change. As temperatures rise, these species are likely to move upstream to find cooler habitats (*6*). The proliferation of small hydropower projects will obstruct their path. Other species may try to move to higher elevations to escape invasive species; the proposed dams will block their access as well (*7*). Moreover, the projects will disrupt the activities of apex predators that depend on headwaters for feeding, breeding, and migration (*8, 9*).

High warming (*10*) and seismicity (*11*) rates in the Himalaya make existing small hydropower infrastructures unstable. During the 2021 Chamoli disaster, for example, an avalanche washed out a small hydropower project and the resulting massive debris flow left 200 people killed or missing, with an estimated loss of US$223 million (*12*). Meanwhile, the Himalaya's rich natural capital supports millions of livelihoods. The small hydropower projects, if built as proposed, would impair ecosystem services and economic growth.

A policy overhaul–particularly for fragile landscapes like Himalaya–is urgently needed. Renewable plans should focus on solar, wind, and biomass strategies instead of small hydropower projects. India has committed to developing without destroying the environment, enhancing climate resilience, and reducing vulnerability to climate change. Only by protecting the Himalaya from further fragmentation can the country live up to its promises.

Aashna Sharma[1], Vineet K. Dubey[1], Yogesh K. Rawal[2], Kuppusamy Sivakumar[3], Jeyaraj Antony Johnson[1]*, Sambandam Sathyakumar[1]
[1]Wildlife Institute of India, Chandrabani, Dehradun 248001, India. [2]Department of Zoology, Panjab University, Chandigarh 160014, India. [3]Department of Ecology and Environmental Sciences, Pondicherry University, Pondicherry 605014, India.
*Corresponding author. Email: jaj@wii.gov.in

REFERENCES AND NOTES

1. Press Information Bureau, Government of India, "National statement by Prime Minister Shri Narendra Modi at COP26 summit in Glasgow" (2021); https://pib.gov.in/PressReleasePage.aspx?PRID=1768712.
2. A. Kumar, "Small hydropower data base for India covering commissioned projects, under construction projects and future sites" (AHEC IIT, Roorkee, 2016).
3. Ministry of Jal Shakti, Department of Water Resources, River Development & Ganga Rejuvenation, Government of India, "National register of large dams" (2018).
4. R. E. Grumbine, M. K. Pandit, *Science* **339**, 36 (2013).
5. K. Chandra *et al*., Eds., *Faunal Diversity of Indian Himalaya* (Zoological Survey of India, Kolkata, 2018).
6. A. Sharma *et al*., *Ecol. Indic.* **120**, 106900 (2021).
7. A. Sharma *et al*., *J. Appl. Ecol.* **58**, 2415 (2021).
8. J. P. Bhatt, M. K. Pandit, *Rev. Fish Biol. Fish.* **26**, 25 (2016).
9. P. S. Jamwal *et al*., *Divers. Distrib.* 10.1111/ddi.13377 (2021).
10. "Climate change 2021: The physical science basis," Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge University Press, 2021).
11. L. Dal Zilio *et al*., *Nat. Rev. Earth Environ.* **2**, 251 (2021).
12. H. Shugar *et al*., *Science* **373**, 300 (2021).

10.1126/science.abp8882

Addressing patent gender disparities

Women comprise 29% of the science, technology, engineering, and mathematics (STEM) workforce but only 12.8% of US patent inventors (*1*). Institutional structures and biases contribute to the patent equity gap (*2–4*). To increase the number of women who hold patents, institutions must give everyone more information about patents, ensure that women receive equal credit for their work, and provide women with equitable access to work opportunities.

Women are less likely than their male counterparts to pursue patents for their inventions because they are less familiar with the patent process and less likely to see themselves as "inventive" (*4*, *5*). Institutions can address these gaps by creating inventor education programs that use equitable language (*6*). For example, to address differences in self-perception between men and women, institutions could refer to the paperwork required to internally disclose an invention and start the patent process as "progress reports" or "creativity disclosures" instead of the standard "inventor disclosure forms." Institutions could also provide office hours and resources to guide new inventors of all genders through the process.

In some cases, an inventor applies for a patent but unintentionally excludes women who collaborated on the project from the disclosure. Prepopulating disclosure forms with the names of relevant past and present laboratory members can

PHOTO: AASHNA SHARMA

help negate gender-related biases and increase inventor disclosure compliance (*7*, *8*). If applicants are prompted to consider all contributors, they are more likely to include inventors whom they might otherwise overlook.

It is possible that women are less likely to develop technology eligible for patents because they have fewer opportunities to achieve their professional goals (*4*). Institutions must recognize and address evolving societal circumstances that disparately affect women. For example, institutions can create more equitable opportunities by providing affordable and accessible childcare to all employees and promoting policies to ensure that administrative tasks are not delegated more frequently to women than to men (*3*, *4*).

Closing the patent equity gap will require data showing how many women and men are applying for patents and whether specific strategies successfully mitigate disparities. All corporate, governmental, and academic institutions should collect and make public data on the gender distribution of all employees compared with the gender representation of patent inventors. Gendered patent disparity should be quantified and tracked over time, and effective diversity, equity, and inclusion programs [e.g., (*9*)] should be implemented and monitored for success. By making these changes, institutions can wield their power to promote gender equity among patent applicants. All inventors deserve to be recognized for their creative work.

Jordana Goodman
Boston University School of Law, Boston, MA 02215, USA. Email: jordi@bu.edu

REFERENCES AND NOTES

1. A. Tulle *et al*., "Progress and potential: 2020 update on US women inventor-patentees" (US Patent and Trademark Office, 2020); www.uspto.gov/sites/default/files/documents/OCE-DH-Progress-Potential-2020.pdf.
2. G. Giasi, in *The Human Rights-Based Approach to Stem Education*, T. Tajmel, K. Starl, S. Spintig, Eds. (Waxmann, 2021), pp. 193–208.
3. J. C. Williams, M. Multhaup, "For women and minorities to get ahead, managers must assign work fairly," *Harvard Business Review* (2018).
4. E. Shaw, H. Mariano, "Tackling the gender and racial patenting gap to drive innovation: Lessons from women's experiences," *Institute for Women's Policy Research* (2021).
5. "2011 Lemelson-MIT Invention Index uncovers inventive perception among young Americans," *Technology and Engineering Teacher* (2011).
6. J. Fishman, in *Linguistic Human Rights: Overcoming Linguistic Discrimination*, T. Skutnabb-Kangas, R. Phillipson, Eds. (Mouton de Gruyter, 1995), p. 58.
7. A. Llorens *et al*., *Neuron* **109**, 2047 (2021).
8. M. A. Fonseca, S. B. Grimshaw, *J. Public Pol. Market.* **36**, 2013 (2017).
9. C. A. Moss-Racusin, *CBE Life Sci Educ*. **15**, ar29 (2016).

10.1126/science.abq5927

Advancing racial equity in Brazil's academia

People in Brazil who identify as Black, including those with major African inheritance and those who belong to mixed-raced ethnic groups (*1*), have historically been underrepresented in postsecondary education as a result of racial and socioeconomic inequities (*2*, *3*). In Brazil, public universities are free but highly competitive, favoring students from wealthier families who generally had access to better, private secondary education (*4*). Black activism in Brazil has highlighted the inadequate access to education and stimulated debates on how to apply antiracist strategies to achieve racial equity (*5*). In 2013, the Labour Party established race-conscious affirmative actions in public universities, which provided aid to low-income Black people, a change that greatly increased the number of Black undergraduate and graduate students (*6*, *7*). However, President Jair Bolsonaro's opposition to race-based affirmative action and hostility toward higher education jeopardize these successful measures (*8*, *9*).

The Bolsonaro administration's budget cuts to science and education limit the funding available for programs that advance diversity in academia by subsidizing housing, food, and transportation for low-income public university students, many of whom are Black. Given the success of affirmative action policies in improving racial equity in Brazilian universities (*6*, *7*), such programs should be maintained and expanded rather than cut. If decision-makers do not reevaluate this budget, Black people and other minorities will likely be pushed out of academia, and Brazil will lose a generation of bright, underrepresented scientists (*9*, *10*).

As Black scholars who lived through the paradigm shift toward racial equality in higher education in Brazil, we express our profound concerns about the dismantlement of diversity initiatives in postsecondary education. We also demand urgent attention from lawmakers to proposed legislation PL 1788/21, currently under consideration by the Chamber of Deputies, that would maintain racial quotas in higher education for 10 more years (*11*). It is urgent to bring this discussion to light; without race-conscious and socioeconomic affirmative actions in the next few years, Brazil will likely face an unprecedented setback in racial equity in academia.

Willams Oliveira[1]*, Bruno E. Soares[2,3], Piatã Marques[2,3], Caroline Souza[4], Arildo S. Dias[5,3], Marianne Bello[6,3], Lis Custódio[7,3]
[1]Programa de Pós-Graduação em Biologia Vegetal, Departamento de Botânica, Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil. [2]Department of Biological Sciences, University of Toronto-Scarborough, Toronto, ON, Canada. [3]Odu Initiative for Black Representativeness in Science, Rio de Janeiro, Rio de Janeiro, Brazil. [4]Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, São Bernardo do Campo, São Paulo, Brazil. [5]Department of Physical Geography, Goethe University, Frankfurt am Main, Hessen, Germany. [6]Programa de Pós-Graduação em Ecologia e Evolução, Departamento de Ecologia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil. [7]Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil.
*Corresponding author.
Email: willams-costa@hotmail.com

REFERENCES AND NOTES

1. L. Lynch, "Race and affirmative action in Brazil," *The Society Pages* (2018).
2. A. Artes, A. M. Ricoldi, *Cadernos de Pesquisa* **45**, 859 (2015).
3. R. R. Valente, *Race Ethnic. Educ*. **20**, 851 (2017).
4. E. Telles, M. Paixão, "Affirmative action in Brazil," *Lasa Forum* **XLIV**, 10 (2013).
5. N. L. Gomes, *Educação e Sociedade* **33**, 727 (2012).
6. Instituto Brasileiro de Geografia e Estatística (IBGE), "Síntese de indicadores sociais: Uma análise das condições de vida da população brasileira" (IBGE, Rio de Janeiro, 2019) [in Portuguese].
7. IBGE, "Desigualdades sociais por cor ou raça no Brasil" (IBGE, Rio de Janeiro, 2019) [in Portuguese].
8. L. M. Diele-Viegas *et al*., *Science* **374**, 948 (2021).
9. C. H. L. Silva-Júnior *et al*., *Science* **374**, 948 (2021).
10. S. L. Costa, S. M. B. Dias, *Jornal de Políticas Educacionais* **9**, 51 (2015) [in Portuguese].
11. Brasil, Câmara dos Deputados, "Projeto de Lei nº 1788/2021, de 12 de maio de 2021" (2021) [in Portuguese].

10.1126/science.abq5212

TECHNICAL COMMENT ABSTRACTS

Comment on "Cryoforged nanotwinned titanium with ultrahigh strength and ductility"

Yazhou Guo, Xiaolei Wu, Qiuming Wei

We analyze the results of Zhao *et al*. (Report, 17 September 2021, p. 1363) with a focus on the mechanical properties and microstructural evolution. We conclude that their results, together with the explanations and interpretations, are confusing, misleading, or even wrong.

Full text: dx.doi.org/10.1126/science.abo3440

Response to Comment on "Cryoforged nanotwinned titanium with ultrahigh strength and ductility"

Shiteng Zhao, Ruopeng Zhang, Qin Yu, Jon Ell, Robert O. Ritchie, Andrew M. Minor

We address the three main points of Guo *et al*. They claim that we should have used the engineering stress versus engineering strain curves to infer the mechanical properties of our nanotwinned titanium, question our sample design on the basis of a finite-element analysis, and doubt the immobility of some preexisting grain/twin boundaries in our electron backscatter diffraction micrographs. We find their analysis to be groundless and to contain many inconsistencies.

Full text: dx.doi.org/10.1126/science.abo5247

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

Dragonflies use visual stimuli to coordinate righting reflexes while flying.

BIOMECHANICS

Following formidable flyers

Flying involves complicated maneuvers, not the least of which is the process of returning to an upright flying position after being flipped over in midair. Wang *et al.* used a combination of experiments and biophysical models to understand this process in dragonflies, which are adept flyers. The authors found that righting involves a series of signals beginning with the visual system through to wing pitch muscles. This approach revealed connections between neural signals and physical processes that could be used to study flight mechanics across species. —SNV

Science, abg0946, this issue p. 754

NEUROSCIENCE

Emotional memory encoding in REM sleep

During rapid eye movement (REM) sleep, emotional memories are consolidated in the prefrontal cortex. However, we are still far from a mechanistic understanding of the processes involved. Aime *et al*. investigated the impact of REM sleep on cellular and subcellular activities, microcircuit connectivity, plasticity, and behavior. The authors quantified how somatic and dendritic activity differ in mice during wakefulness, REM sleep, and non-REM sleep, and probed how interneurons cause these differences. They further identified a specific area in the thalamus that activates a subpopulation of interneurons that decouple the dendrite from the somatic output. Finally, by optogenetic intervention during REM sleep, learning of discrimination between danger and safety could be enhanced or diminished. These findings provide a better understanding of the processing of emotions during sleep. —PRS

Science, abk2734, this issue p. 724

SOLAR CELLS

Large-area tandem perovskite solar cells

Tandem solar cells allow more of the solar spectrum to be used. For all-perovskite implementations with large illumination areas, different bandgap compositions must be grown with fully scalable methods. Xiao *et al*. blade coated high-quality, wide-bandgap perovskite layers by tuning the cesium concentration in a mixed solvent system. They avoided diffusion between the perovskite layers with a tin oxide layer grown by atomic layer deposition that also served as an electron extractor. Small-area cells (1 square centimeter) had a power conversion efficiency (PCE) of about 25%, and a 20-square-centimeter module had a certified PCE of 21.7%. The encapsulated tandem module maintained more than 75% of its initial PCE after maximum power point operation for over 500 hours in ambient air. —PDS

Science, abn7696, this issue p. 762

FERROELECTRICS

Ion enhanced polarization

Hafnium oxide is an exciting material because it has ferroelectric behavior that makes it attractive for various device applications. Kang *et al*. found that the ferroelectric properties improve by bombarding films of hafnium oxide with a beam of helium ions. The ion bombardment creates oxygen vacancies and strain changes from helium implantation that push more of the polycrystalline samples into the ferroelectric orthorhombic phase. This method may become an important tool for stabilizing the ferroelectric phase for the next generation of electronic devices. —BG

Science, abk3195, this issue p. 731

ORGANIC CHEMISTRY

Boron flips the script on cyclohexane

Conformational analysis of disubstituted cyclohexanes played an outsized role in the development of organic chemistry. It is striking in that context that the comparatively less stable isomers remain challenging to access. Li *et al*. report a versatile method to prepare 1,2-*cis*, 1,3-*trans*, and 1,4-*cis*-disubstituted cyclohexanes, each of which requires a substituent to adopt the disfavored axial orientation. The method relies on a bulky boron substituent one carbon away from the ring to bias the approach of a nickel catalyst. —JSY

Science, abn9124, this issue p. 749

PHOTO: SANDER MEERTINS/ISTOCK.COM

PLANETARY SCIENCE

An extended history of water on mars

Early Mars was wet and relatively warm, but Mars is much colder today, and water there mostly occurs in the form of ice. When this transition occurred is the focus of much Mars research. Recent remote-sensing evidence has suggested that liquid water persisted into the past 2.9 billion years of Mars history. Liu *et al.* reinforce this argument using spectroscopic data obtained by China's Zhurong rover. These data include the first surface detection of hydrous minerals in a sedimentary rock from the Utopia Planitia region of Mars. They also indicate the presence of liquid surface water during sediment deposition in the area, which likely occurred some time in the past billion years. —KVH

Sci. Adv. 10.1126/sciadv.abn8555 (2022).

WETLAND ECOLOGY

Global shifts in tidal wetlands

Ecologically and economically important coastal wetlands are threatened by sea level rise and land use change. Murray *et al*. used high-resolution satellite imagery to assess the global extent of tidal wetlands and changes in wetland extent and distribution over the past two decades. They found that although over 13,000 square kilometers of tidal wetland have recently been lost, much of this decreasing extent has been offset by the creation of new wetlands. The greatest losses and gains were in tidal flats, but mangrove ecosystems showed the largest net decline in area globally. Direct human impacts on wetlands, including land transformation and restoration, are detectable from satellite imagery and account for 27% of wetland losses and gains. —BEL

Science, abm9583, this issue p. 744

PAIN

Beneficial inflammation

The mechanisms mediating the transition from acute to chronic pain are not well understood. Analyzing samples from patients and animal models, Parisien *et al*. found that neutrophil activation–dependent inflammatory genes were up-regulated in subjects with resolved pain, whereas no changes were observed in patients with persistent pain. In rodents, anti-inflammatory treatments prolonged pain duration, and this effect was abolished by neutrophil administration. Clinical data also showed that use of anti-inflammatory drugs was associated with increased risk of persistent pain, suggesting that anti-inflammatory treatments might have negative effects on pain duration. —MM

Sci. Transl. Med. **14**, eabj9954 (2022).

Tidal wetlands, such as this salt marsh in Schiermonnikoog, Netherlands, are dynamic habitats experiencing growth in some locations but greater net loss globally.

PHOTOS (LEFT TO RIGHT): TASFOTO/ALAMY STOCK PHOTO; CITIZEN OF THE PLANET/EDUCATION IMAGES/UNIVERSAL IMAGES GROUP/GETTY IMAGES

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

SOCIAL SCIENCES

Neighborhood pollution discriminates

One of the strongest predictors of health in the United States is zip (postal) code because social and environmental conditions can vary so much by location. The health status of poor and ethnic minority neighborhoods is not only compromised by a lack of resources but also by higher air pollution levels in immediate surroundings and adjacent neighborhoods. However, it is unclear whether pollution exposure changes when people travel to nonadjacent neighborhoods for employment, errands, or social engagements. Brazil used mobile phone location data collected across the 100 largest US cities and found that Hispanic neighborhoods (whether poor or more affluent) were exposed to the highest pollution levels and residents tended to travel to nonadjacent regions also experiencing high pollution. By contrast, more affluent, predominantly white neighborhoods were least exposed to air pollution. Therefore, minorities can carry a previously unmeasured additional burden of health disadvantage across their social networks. —EEU *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117776119 (2022).

Residents of poor and ethnic minority neighborhoods tend to be exposed to unacceptable levels of air pollution at all points of their social networks.

CANCER

A microbial oncometabolite

The gut microbiota have been associated with various types of cancer. For example, particular bacterial species may contribute to the initiation and progression of colorectal cancer (CRC). *Fusobacterium nucleatum* has been identified as a pro-CRC bacterium, but the underlying mechanism is unknown. Ternes *et al*. confirmed that *F. nucleatum* was abundant in stool samples from 63 patients with CRC. In coculture experiments, they

ANTHROPOLOGY

Human resilience to climate change

It is difficult to determine how past human populations coped with climate change. Cultural transformations and climate events have certainly been correlated by historians. Plunkett and Swindles challenge the view that human societies are necessarily vulnerable to collapse in the face of climate change. A marginal (cool "mossy bog") upland landscape in Northern Ireland has been continuously occupied since the Middle Ages despite climate oscillation. A peat core sample supplied detailed high-resolution climate proxy data from volcanic ash, fossil, and pollen deposits. These data, coupled with records of plague, war, and famines, show that none of these events interrupted the mixed agricultural subsistence economy of this society for long. Either the population was sufficiently isolated to escape the consequences of these events or robust social organization allowed them to recover. Either way, this detailed record from western Europe provides an indication of the adaptability of some human social organizations during environmental stress. —CA *PLOS ONE* **17**, e0266680 (2022).

Despite climate oscillations, wars, and famines, the peat uplands of Slieveanorra, Northern Ireland, have been continuously occupied by humans for over 1000 years.

found that *F. nucleatum* altered the signaling and metabolism of cells from patient CRC, and this was caused by the increased secretion of formate by *F. nucleatum*. Furthermore, administration of formate increased CRC tumor size and metastasis in vivo, indicating that formate is an oncometabolite that drives CRC progression. —GKA

Nat. Metab. **4**, 458 (2022).

SYNTHETIC BIOLOGY

Bacteria engineered to protect the gut

Although treatment with broad-spectrum antibiotics provides invaluable protection from infections, it also can also disrupt microbial communities in the gut. Cubillos-Ruiz *et al.* propose a strategy in which treated animals are fed bacteria that produce and secrete an enzyme to degrade β-lactam antibiotics in the gut. This tactic allows the antibiotic to reach effective concentrations in the serum but decreases disruption of the gut microbiota. Critically, the engineered bacteria are not themselves resistant to antibiotic, and the two-gene biosynthetic strategy used is unlikely to allow transfer of resistance. —LBR

Nat. Biomed. Eng. 10.1038/s41551-022-00871-9 (2022).

PHYSICS

Thermal Hall effect and phonons

The material α-$RuCl_3$ has garnered much interest as a potential realization of the so-called Kitaev spin liquid. A key piece of evidence has come from the observation of half-quantization of the thermal Hall conductivity, which is consistent with the predictions of the Kitaev model. However, these measurements show strong sample dependence, which complicates interpretation. Lefrançois *et al.* measured thermal Hall conductivity as well as longitudinal thermal conductivity in five different samples of α-$RuCl_3$. Based on a striking similarity between the temperature dependence of the phonon-dominated longitudinal thermal conductivity and thermal Hall conductivity, the researchers concluded that the thermal Hall effect likely has a large contribution from phonons, which should be taken into account when interpreting similar experiments. —JS

Phys. Rev. X **12**, 021025 (2022).

QUANTUM NETWORKS

A dynamic quantum memory

Because they are robust and fast, photons are ideal carriers of quantum information in long-distance quantum networks. For a practical communication platform, the photons also need to be manipulated at the local level at nodes or exchange points across the network, requiring information to be reliably encoded, stored, and the information retrieved from the photonic qubits. Craiciu *et al.* demonstrate such functionality using an ensemble of erbium ions embedded within in a solid-state matrix of yttrium orthosilicate. With the ions operating as a dynamic quantum memory for photons at communication wavelengths and the approach being compatible with integrated photonics, the results are promising for the possible integration into large-scale optical fiber networks. —ISO

Optica **8**, 114 (2021).

MARINE HEATWAVES

The long and the short of it

Continued global warming has increased the frequency and severity of marine heatwaves. These events, of finite duration like their atmospheric counterparts, can affect both the physical and biological environments of the regions in which they occur. Mogen *et al.* studied the first year of the Northeast Pacific marine heat wave of 2013 to 2016, popularly referred to as the "Blob," and found that although it temporarily mitigated the effects of ocean acidification, because of the changes that it induced in ocean circulation and mixing, it also reduced ocean oxygen content through warming. These results are early steps toward a better understanding of the potential impacts of multiple stressors that may occur during marine heat waves. —HJS

Geophys. Res. Lett. 10.1029/2021GL096938 (2022).

PHOTO: BALLYGALLY VIEW IMAGES/SHUTTERSTOCK

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

HUMAN CELL ATLAS

Mapping cell types in the human body

The Tabula Sapiens is a molecular reference atlas for more than 400 cell types of the human body. The Tabula Sapiens Consortium used single-cell transcriptomics to measure the messenger RNA molecules in each of nearly 500,000 cells from 24 tissues and organs (see the Perspective by Liu and Zhang). These data enable new insights into how the human genome parts list is used to create distinct cell types within the human organism. In addition to creating a detailed molecular definition of these cell types, the atlas reveals many other aspects of human biology, including how the same gene can be spliced differently in different cell types, how shared cell types in different tissues can have subtle differences in their identities, and how clones of the immune system can be shared across tissues. —LMZ and DJ

Science, abl4896, this issue p. 711; see also abq2116, p. 695

HUMAN CELL ATLAS

Cartography of human cells

The function of disease genes active in different cell types is modulated to meet the needs of the different tissues and organs in which the cells reside. Resolving these differences is critical to understanding homeostasis and disease. However, single-cell atlases generated to date have largely focused on individual tissues. Eraslan *et al*. applied single-nucleus RNA sequencing to frozen, banked samples from eight healthy human organs from 16 donors and characterized cell populations across tissues, including tissue-resident myeloid and fibroblast populations, and their role in tissue support and immunity (see the Perspective by Liu and Zhang). Using this cross-tissue atlas, the authors linked specific cell populations to monogenic and polygenic diseases, suggesting cell- and tissue-specific programs. —LMZ and DJ

Science, abl4290, this issue p. 712; see also abq2116, p. 695

HUMAN CELL ATLAS

Immune cell diversity in the human body

The human immune system is composed of many different cell types spread across the entire body, but little is currently known about the fine-grained variations in these cell types across organs. Using single-cell genomics, Domínguez Conde *et al*. examined the gene expression profile of more than 300,000 individual immune cells extracted from 16 different tissues in 12 deceased adult organ donors (see the Perspective by Liu and Zhang). Cell identity was assigned using CellTypist, an automated cell classification tool designed by the authors. In-depth data analysis revealed insights into how the immune system adapts to function effectively in different organ contexts. —LMZ and DJ

Science, abl5197, this issue p. 713; see also abq2116, p. 695

MICROBIOTA

Gut bacteria go on the record

Bacteria in the gut dynamically respond to each another, as well as to their host's diet and immune system. However, current approaches to measuring changes in gene expression in this setting have been limited and are generally invasive. Schmidt *et al*. use a CRISPR-based recording method (Record-seq) to capture the transcriptional changes that occur in *Escherichia coli* bacteria as they pass through the intestines (see the Perspective by Zahavi and Segal). Deep sequencing of these DNA recordings revealed characteristic signatures based on dietary, host, and microbiota contents at different levels of the intestine. This work provides additional perspectives on how diet, inflammation, and microbial interactions in the body shape the health of mammalian hosts. —STS

Science, abm6038, this issue p. 714; see also abq1455, p. 697

NANOCHEMISTRY

Fluorine for faster water transport

Typically, as the diameter of a pipe or channel is shrunk, the flow rate per cross-sectional area will decrease because of the greater surface area to volume ratio. Itoh *et al*. designed a series of fluorous oligoamide nanorings that undergo supramolecular polymerization to form nanochannels with dense packing of organofluorine groups (see the Perspective by Shen). The intense electronegativity disrupts the formation of water clusters, so that individual water molecules flow through the smallest channels faster than through larger ones. Chloride ions are strongly repelled and cannot travel through the channels. —MSL

Science, abd0966, this issue p. 738; see also abo2953, p. 698

FOREST ECOLOGY

Uncoupled carbon uptake and storage

Forests are expected to help mitigate climate change by sequestering carbon, with elevated carbon dioxide boosting photosynthesis and carbon uptake. However, the amount of carbon that can be stored in wood also depends on temperature, water, and nutrient availability. Cabon *et al*. examined temporal correlations between trees' carbon uptake and woody growth by combining data on tree rings and gross primary productivity measures from 78 forests with carbon dioxide flux towers (see the Perspective by Green and Keenan). They found weak correlations between productivity and woody growth, which responded differently to seasonal temperatures and water availability. Their work shows that limits to tree growth, particularly in dry and cold areas, may constrain potential forest carbon storage. —BEL

Science, abm4875, this issue p. 758; see also abo6547, p. 692

VACCINES

MAdCAM-1 about aging

As people age, their immune responses to vaccines and infections diminish, increasing the risk for serious illness. These diminished immune responses have been tied to poor germinal center responses in older people, yet the exact mechanism for how this effect occurs is unclear. Denton *et al*. vaccinated adult or aged mice with a model immunogenic protein to track germinal center responses overtime. They found that the microenvironment of the lymph node was impaired in older mice, as defined by a poor response of MAdCAM-1$^+$ stromal cells to vaccination. These cells were partially rescued with a Toll-like receptor 4 (TLR4) agonist, which improved the initiation of germinal center immune responses. Together, these data suggest that targeting TLR4 might improve the efficacy of vaccination in older people. —DAE

Sci. Immunol. **7**, eabk0018 (2022).

NEUROSCIENCE

Bridging adhesion signaling for learning

Biliverdin reductase (BVR) and its product, bilirubin, have antioxidant functions in the brain. Vasavda *et al*. found an

additional synaptic role for BVR that was independent of either bilirubin or biliverdin. BVR linked focal adhesion kinases to an effector kinase at synapses in the mouse hippocampus, which enabled signaling between these kinases and synaptic plasticity, a process that underlies learning in which synaptic strength is calibrated in response to changing inputs. Mice lacking BVR had impaired focal adhesion signaling and showed performance deficits on cognitive tests. —LKF

Sci. Signal. **15**, abh3066 (2022).

BATTERIES

Theory-guided materials research

A reliable energy storage ecosystem with high energy density and long lifetime storage is imperative for a renewable energy future. However, the further development of rechargeable batteries lacks a better theoretical and experimental understanding of electrochemical mechanisms and structure-property relationships. In a Review, Eng *et al.* discuss the interplay between theory and experiment in battery materials research, the combination of which will contribute to new material predictions and investigations of mechanisms. The authors focus on specific examples in state-of-the-art lithium-ion, lithium-metal, sodium-metal, and all-solid-state batteries. This theory-experiment framework can also be extended to multivalent batteries. Interdisciplinary collaboration is expected to provide universal principles to develop next-generation rechargeable batteries. —XWDL

Sci. Adv. 10.1126/sciadv.abm2422 (2022).

NEUROIMMUNOLOGY

T cells in the elderly touch a nerve

With age, the ability to bounce back from nervous system injuries becomes increasingly limited. However, the molecular and cellular mechanisms underlying aging-dependent regenerative decline remain poorly understood. Zhou *et al*. used RNA sequencing and imaging to study dorsal root ganglia in a mouse model of sciatic nerve injury (see the Perspective by Gaudenzio and Liblau). Compared with younger mice, injured DRG sensory neurons in older mice expressed higher levels of the chemokine CXCL13, which attracted $CXR5^{+}CD8^{+}$ T cells. T cell–neuron cross-talk activated caspase 3, which impaired pAKT, and pS6 signaling causing regenerative failure. When caspase 3, CXCR5, or CXCL13 was therapeutically targeted, the authors were able to restore neuronal regeneration and recovery in older mice to levels comparable to their younger counterparts. —STS

Science, abd5926, this issue p. 715; see also abp9878, p. 694

QUANTUM SIMULATION

Hydrodynamics of spin chains

Computing the dynamics of quantum many-body systems is notoriously difficult. Nevertheless, at long times after a quench, even these complicated systems are predicted to feature hydrodynamic transport properties. Two groups have used widely tunable and exquisitely controllable atomic systems to study the dynamics of spin propagation in quantum magnetic chains (see the Perspeective by Morningstar and Bakr). Joshi *et al*. varied the range of interactions in a chain of calcium ions and observed emergent hydrodynamics in several different universality classes. Wei *et al*. used a quantum gas microscope to monitor spin transport in a Heisenberg chain of cold atoms and found that the system exhibited the so-called Kardar-Parisi-Zhang superdiffusion. —JS

Science, abk2397, abk2400, this issue p. 716, p. 720; see also abn6376, p. 699

RESEARCH ARTICLE SUMMARY

HUMAN CELL ATLAS

The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans

The Tabula Sapiens Consortium*

INTRODUCTION: Although the genome is often called the blueprint of an organism, it is perhaps more accurate to describe it as a parts list composed of the various genes that may or may not be used in the different cell types of a multicellular organism. Although nearly every cell in the body has essentially the same genome, each cell type makes different use of that genome and expresses a subset of all possible genes. This has motivated efforts to characterize the molecular composition of various cell types within humans and multiple model organisms, both by transcriptional and proteomic approaches. We created a human reference atlas comprising nearly 500,000 cells from 24 different tissues and organs, many from the same donor. This atlas enabled molecular characterization of more than 400 cell types, their distribution across tissues, and tissue-specific variation in gene expression.

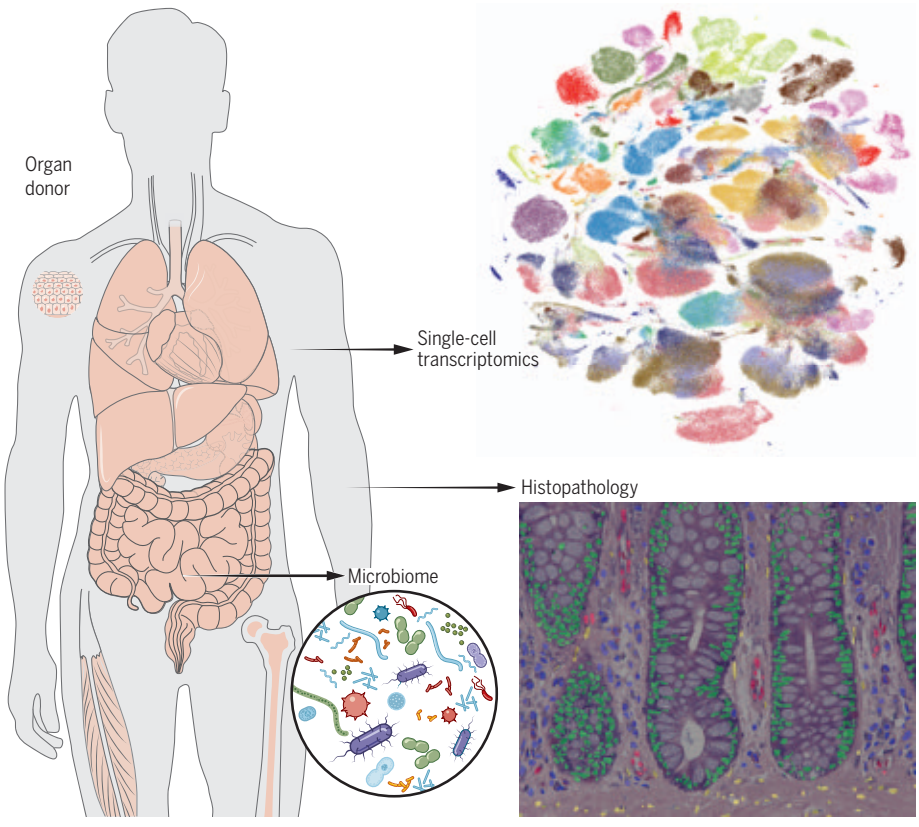
RATIONALE: One caveat to current approaches to make cell atlases is that individual organs are often collected at different locations, collected from different donors, and processed using different protocols. Controlled comparisons of cell types between different tissues and organs are especially difficult when donors differ in genetic background, age, environmental exposure, and epigenetic effects. To address this, we developed an approach to analyzing large numbers of organs from the same individual.

RESULTS: We collected multiple tissues from individual human donors and performed coordinated single-cell transcriptome analyses on live cells. The donors come from a range of ethnicities, are balanced by gender, have a mean age of 51 years, and have a variety of medical backgrounds. Tissue experts used a defined cell ontology terminology to annotate cell types consistently across the different tissues, leading to a total of 475 distinct cell types with reference transcriptome profiles. The full dataset can be explored online with the cellxgene tool.

Data were collected for the bladder, blood, bone marrow, eye, fat, heart, kidney, large intestine, liver, lung, lymph node, mammary, muscle, pancreas, prostate, salivary gland, skin, small intestine, spleen, thymus, tongue, trachea, uterus, and vasculature. Fifty-nine separate specimens in total were collected, processed, and analyzed, and 483,152 cells passed quality control filtering. On a per-compartment basis, the dataset includes 264,824 immune cells, 104,148 epithelial cells, 31,691 endothelial cells, and 82,478 stromal cells. Working with live cells, as opposed to isolated nuclei, ensured that the dataset includes all mRNA transcripts within the cell, including transcripts that have been processed by the cell's splicing machinery, thereby enabling insight into variation in alternative splicing.

The Tabula Sapiens also provided an opportunity to densely and directly sample the human microbiome throughout the gastrointestinal tract. The intestines from two donors were sectioned into five regions: the duodenum, jejunum, ileum, and ascending and sigmoid colon. Each section was transected, and three to nine samples were collected from each location, followed by amplification and sequencing of the 16*S* ribosomal RNA gene.

CONCLUSION: The Tabula Sapiens has revealed discoveries relating to shared behavior and subtle, organ-specific differences across cell types. We found T cell clones shared between organs and characterized organ-dependent hypermutation rates among B cells. Endothelial cells and macrophages are shared across tissues, often showing subtle but clear differences in gene expression. We found an unexpectedly large and diverse amount of cell type–specific RNA splice variant usage and discovered and validated many previously undefined splices. The intestinal microbiome was revealed to have nonuniform species distributions down to the 3-inch (7.62-cm) length scale. These are but a few examples of how the Tabula Sapiens represents a broadly useful reference to deeply understand and explore human biology at cellular resolution. ■



Overview of Tabula Sapiens. Molecular characterization of cell types using single-cell transcriptome sequencing is revolutionizing cell biology and enabling new insights into the physiology of human organs. We created a human reference atlas comprising nearly 500,000 cells from 24 different tissues and organs, many from the same donor. This multimodal atlas enabled molecular characterization of more than 400 cell types.

The complete list of authors and their affiliations is available in the full article online.
*Corresponding author: Stephen R. Quake (steve@quake-lab.org)
Cite this article as The Tabula Sapiens Consortium, *Science* **376**, eabl4896 (2022). DOI: 10.1126/science.abl4896

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl4896

RESEARCH ARTICLE

HUMAN CELL ATLAS

The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans

The Tabula Sapiens Consortium*

Molecular characterization of cell types using single-cell transcriptome sequencing is revolutionizing cell biology and enabling new insights into the physiology of human organs. We created a human reference atlas comprising nearly 500,000 cells from 24 different tissues and organs, many from the same donor. This atlas enabled molecular characterization of more than 400 cell types, their distribution across tissues, and tissue-specific variation in gene expression. Using multiple tissues from a single donor enabled identification of the clonal distribution of T cells between tissues, identification of the tissue-specific mutation rate in B cells, and analysis of the cell cycle state and proliferative potential of shared cell types across tissues. Cell type–specific RNA splicing was discovered and analyzed across tissues within an individual.

Although the genome is often called the blueprint of an organism, it is perhaps more accurate to describe it as a parts list composed of the various genes that may or may not be used in the different cell types of a multicellular organism. Although nearly every cell in the body has essentially the same genome, each cell type makes different use of that genome and expresses a subset of all possible genes (*1*). Therefore, the genome in and of itself does not provide an understanding of the molecular complexity of the various cell types of that organism. This has motivated efforts to characterize the molecular composition of various cell types within humans and multiple model organisms, both by transcriptional (*2*) and proteomic (*3*, *4*) approaches.

Although such efforts are yielding insights (*5*–*7*), one caveat to current approaches is that individual organs are often collected at different locations or from different donors (*8*), are processed using different protocols, or lack replicate data (*9*). Controlled comparisons of cell types between different tissues and organs are especially difficult when donors differ in genetic background, age, environmental exposure, and epigenetic effects. To address this, we developed an approach to analyze large numbers of organs from the same individual (*10*), which we originally used to characterize age-related changes in gene expression in various cell types in the mouse (*11*).

Data collection and cell type representation

We collected multiple tissues from individual human donors (designated TSP1 to TSP15) and performed coordinated single-cell transcriptome analysis on live cells (*12*). We collected 17 tissues from one donor, 14 tissues from a second donor, and five tissues from two other donors (Fig. 1). We also collected smaller numbers of tissues from a further 11 donors, creating biological replicates for nearly all tissues. The donors come from a range of ethnicities, are balanced by gender, have a mean age of 51 years, and have a variety of medical backgrounds (table S1). Single-cell transcriptome sequencing was performed with both fluorescence-activated cell sorting (FACS)–sorted cells in well plates with smart-seq2 amplification as well as 10× microfluidic droplet capture and amplification for each tissue (fig. S1). Tissue experts used a defined cell ontology terminology to annotate cell types consistently across the different tissues (*13*), leading to a total of 475 distinct cell types with reference transcriptome profiles (tables S2 and S3). The full dataset can be explored online with the cellxgene tool through the Tabula Sapiens data portal (*14*).

Data were collected for the bladder, blood, bone marrow, eye, fat, heart, kidney, large intestine, liver, lung, lymph node, mammary, muscle, pancreas, prostate, salivary gland, skin, small intestine, spleen, thymus, tongue, trachea, uterus, and vasculature. Fifty-nine separate specimens in total were collected, processed, and analyzed, and 483,152 cells passed quality control (QC) filtering (figs. S2 to S7 and table S2). On a per-compartment basis, the dataset includes 264,824 immune cells, 104,148 epithelial cells, 31,691 endothelial cells, and 82,478 stromal cells. Working with live cells as opposed to isolated nuclei ensured that the dataset includes all mRNA transcripts within the cell, including transcripts that have been processed by the cell's splicing machinery, thereby enabling insights into variations in alternative splicing.

To characterize the relationship between transcriptome data and conventional histologic analysis, a team of pathologists analyzed hematoxylin and eosin (H&E)–stained sections prepared from nine tissues from donor TSP2 and 13 tissues from donor TSP14 (*14*). Cells were identified by morphology and classified broadly into epithelial, endothelial, immune, and stromal compartments as well as rarely detected peripheral nervous system (PNS) cell types. (Fig. 2A). These classifications were used to estimate the relative abundances of cell types across the four compartments and to estimate the uncertainties in these abundances resulting from spatial heterogeneity of each tissue type (Fig. 2B and fig. S8). We compared the histologically determined abundances with those obtained by single-cell sequencing (fig. S9). Although, as expected, there can be substantial variation between the abundances determined by these methods, in aggregate, we observed broad concordance over a large range of tissues and relative abundances. This approach enables an estimate of true cell type proportions because not every cell type survives dissociation with equal efficiency (*15*). For several of the tissues, we also performed literature searches and collected tables of prior knowledge of cell type identity and abundance within those tissues (table S4). We compared literature values with our experimentally observed frequencies for three well-annotated tissues: the lung, muscle, and bladder (fig. S10).

Immune cells: Variation in gene expression across tissues and a shared lineage history

The Tabula Sapiens can be used to study differences in the gene expression programs and lineage histories of cell types that are shared across tissues. We analyzed tissue differences in the 36,475 macrophages distributed among 20 tissues because tissue-resident macrophages are known to carry out specialized functions (*16*). These shared and orthogonal signatures are summarized in a correlation map (fig. S11A). For example, macrophages in the spleen were different from most other macrophages, and this was driven largely by higher expression of CD5L, a regulator of lipid synthesis (fig. S11B). We also observed a shared signature of elevated epiregulin (EREG) expression in solid tissues, such as the skin, uterus, and mammary, compared with circulatory tissues (fig. S11B).

We characterized lineage relationships between T cells by assembling the T cell receptor sequences from donor TSP2. Multiple T cell lineages were distributed across various tissues in the body, and we mapped their relationships (Fig. 3A). Large clones often reside in multiple organs, and several clones of mucosal-associated invariant T cells are shared across donors (fig. S11C); these cells had characteristic expression of *TRAV1-2* because they are thought to be innate-like effector cells (*17*). Lineage information can also reveal tissue-specific

All authors and their affiliations appear at the end of this paper.
*Corresponding author: Stephen R. Quake (steve@quake-lab.org)

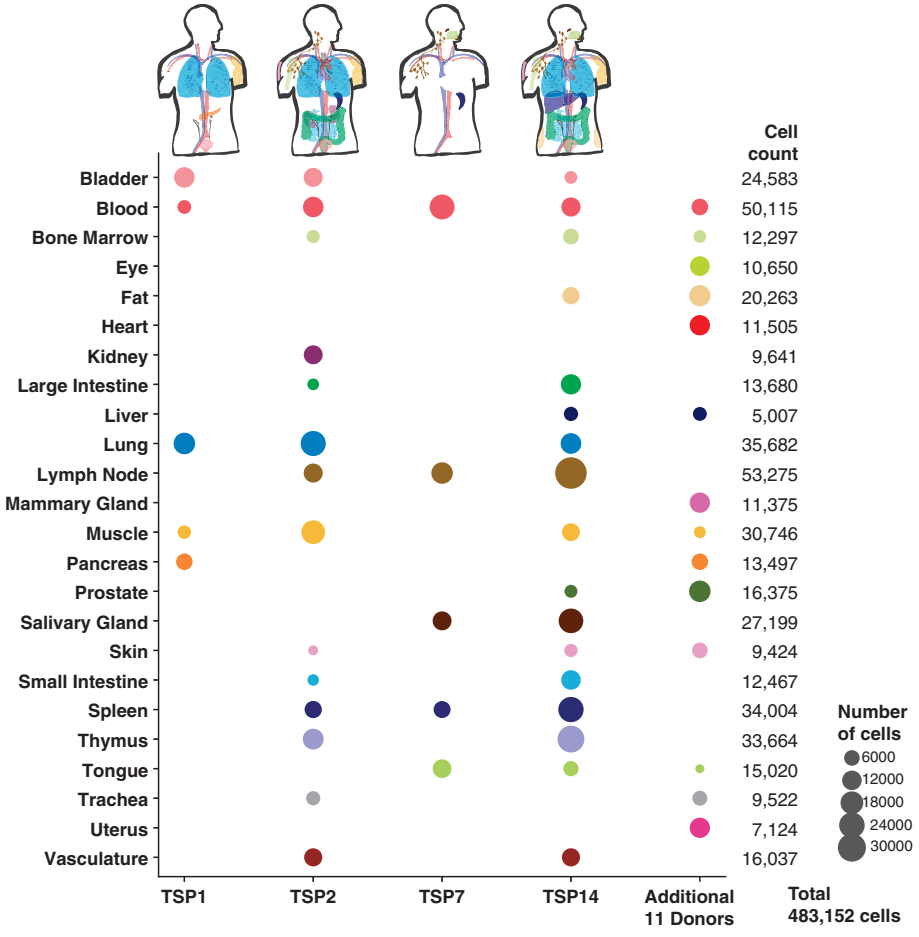


Fig. 1. Overview of Tabula Sapiens. The Tabula Sapiens was constructed with data from 15 human donors; for detailed information on which tissues were examined for each donor, please refer to table S2. Demographic and clinical information about each donor is listed in the supplementary materials and methods and in table S1. Donors 1, 2, 7, and 14 contributed the largest number of tissues each, and the number of cells from each tissue is indicated by the size of each circle. Tissue contributions from additional donors who contributed single or small numbers of tissues are shown in the additional 11 donors column, and the total number of cells for each organ are shown in the final column on the right.

somatic hypermutation rates in B cells. We assembled the B cell receptor sequences from donor TSP2 and inferred the germline ancestor of each cell. The mutational load varies markedly by tissue of residence, with blood having the lowest mutational load compared with solid tissues (fig. S11D). Solid tissues have an order of magnitude more mutations per nucleotide (mean = 0.076; SD = 0.026) compared with the blood (mean = 0.0069), which suggests that the immune infiltrates of solid tissues are dominated by mature B cells.

B cells also undergo class-switch recombination that diversifies the humoral immune response by using constant region genes with distinct roles in immunity. We classified every B cell in the dataset as immunoglobulin A (IgA)–, IgG-, or IgM-expressing and then calculated the relative amounts of each cellular isotype in each tissue (Fig. 3B and table S5). Secretory IgA is known to interact with pathogens and commensals at the mucosae, IgG is often involved in direct neutralization of pathogens, and IgM is typically expressed in naïve B cells or is secreted in the first response to pathogens. Consistent with this, our analysis revealed opposing gradients of prevalence of IgA- and IgM-expressing B cells across the tissues, with blood having the lowest relative abundance of IgA-producing cells and the large intestine having the highest relative abundance (and the converse for IgM-expressing B cells) (Fig. 3B).

Endothelial cell subtypes with tissue-specific gene expression programs

As another example of analyzing shared cell types across organs, we focused on endothelial cells (ECs). Although ECs are often categorized as a single cell type, they exhibit differences in morphology, structure, and immunomodulatory and metabolic phenotypes depending on their tissue of origin. We discovered that tissue specificity is also reflected in their transcriptomes because ECs mainly cluster by tissue of origin (table S6). Uniform manifold approximation and projection (UMAP) analysis (fig. S12A) revealed that the lung, heart, uterus, liver, pancreas, fat, and muscle ECs exhibited the most-distinct transcriptional signatures, reflecting their highly specialized roles. These distributions were conserved across donors (fig. S12B).

Notably, ECs from the thymus, vasculature, prostate, and eye were similarly distributed across several clusters, which suggests not only similarity in transcriptional profiles but in their sources of heterogeneity. Differential gene expression analysis between ECs from these 16 tissues revealed several canonical and previously undescribed tissue-specific vascular markers (Fig. 3C). These data recapitulate tissue-specific vascular markers, such as LCN1 (tear lipocalin) in the eye, ABCG2 (transporter at the blood-testis barrier) in the prostate, and OIT3 (oncoprotein-induced transcript 3) in the liver. Of the potential previously undescribed markers determined by this analysis, SLC14A1 (solute carrier family 14 member 1) appears to be a specific marker for endothelial cells in the heart, whose expression was independently validated with data from the Human Protein Atlas (*18*) (fig. S13).

Lung ECs formed two distinct populations, which is in line with the aerocyte (aCap - EDNRB+) and general capillary (gCap - PLVAP+) cells described in the mouse and human lung (*19*) (fig. S12, C and D). The transcriptional profile of gCaps were also more similar to ECs from other tissues, indicative of their general vascular functions in contrast to the more specialized aCap populations. Lastly, we detected two distinct populations of ECs in the muscle, including a MSX1+ population with strong angiogenic and endothelial cell proliferation signatures and a CYP1B1+ population enriched in metabolic genes, which suggests the presence of functional specialization in the muscle vasculature (fig. S12, E and F).

Alternative splice variants are cell type specific

We used SICILIAN (*20*) to identify alternative splice junctions in Tabula Sapiens using both 10× and smart-seq2 sequencing technologies and found a total of 955,785 junctions (fig. S14, A to E, and table S7). Of these, 217,855 were previously annotated, so our data provide independent validation of 61% of the total junctions cataloged in the entire RefSeq database. Although annotated junctions made up only 22.8% of the total junctions, they represent 93% of total reads, which indicates that previously annotated junctions tend to be expressed at higher levels than previously

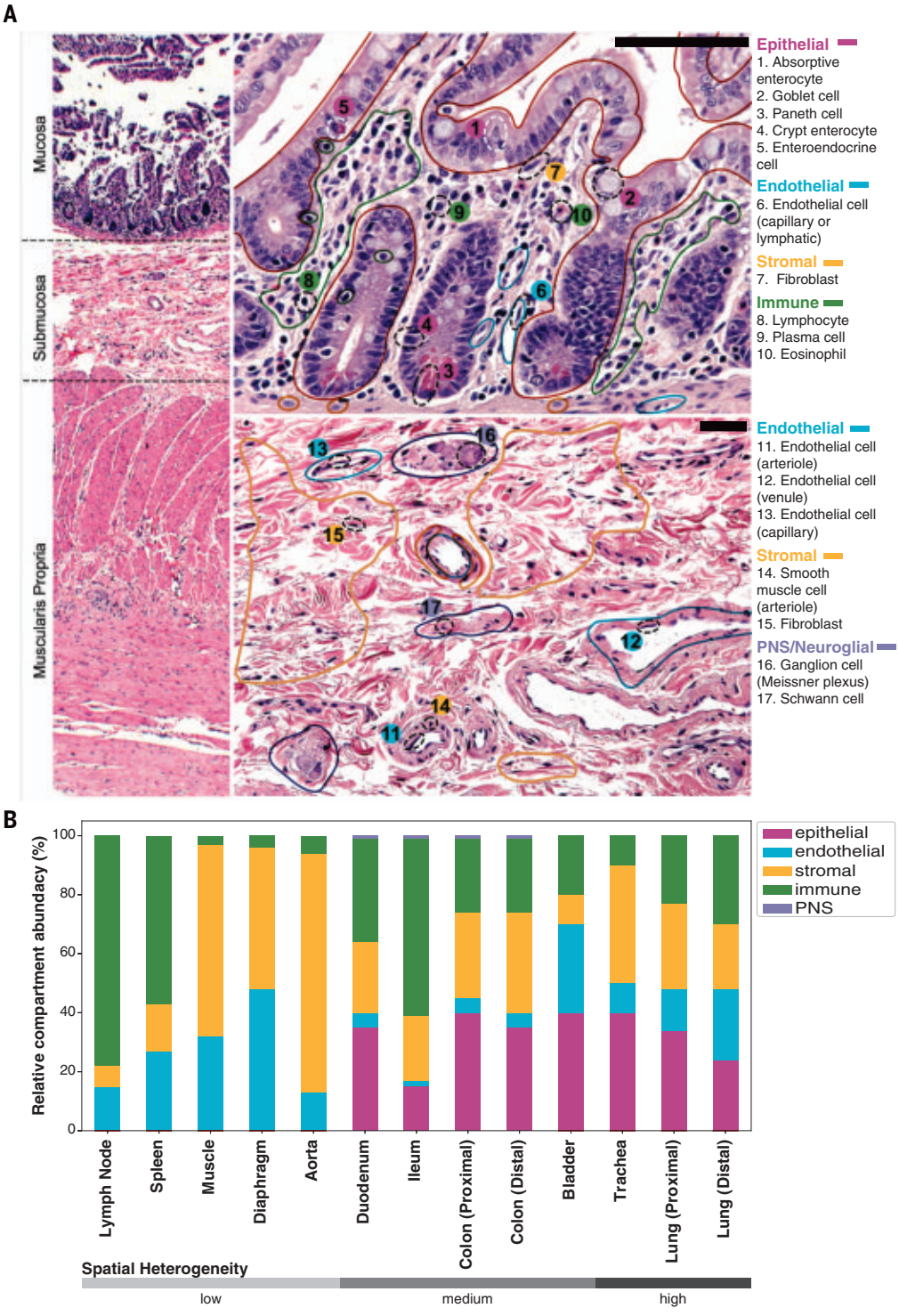


Fig. 2. Comparison of single-cell transcriptomics with conventional histology. Clinical pathology was performed on nine tissues from donors TSP2 and TSP14. (**A**) H&E–stained image used for histology of the colon from TSP2, with compartments (solid colored lines) and individual cell types (dashed black ellipses) identified by the pathologists. (**B**) Coarse cell type representation of TSP2 as morphologically estimated by pathologists across several tissues, ordered by increasing heterogeneity of the tissue. Compartment colors are consistent between (A) and (B).

unidentified junctions. We additionally found 34,624 junctions between previously annotated 3′ and 5′ splice sites (3.6%). We identified 119,276 junctions between a previously annotated site and a previously unannotated site in the gene (12.4%). This leaves 584,030 putative junctions for which both splice sites were previously unannotated—i.e., ~61% of the total detected junctions. Most of these have at least one end in a known gene (94.7%), whereas the remainder represent potential previously undescribed splice variants from unannotated regions (5.3%). In the absence of independent validation, we conservatively characterized all of the unannotated splices as putative previously unknown junctions. We then used the GTEx database (*21*) to seek independent corroborating evidence of these putative junctions and found that reads corresponding to nearly one-third of these previously unknown junctions can be found within the GTEx data (table S7); this corresponds to >300,000 previously undefined validated splice variants revealed by the Tabula Sapiens.

Hundreds of splice variants are used in a highly cell type–specific fashion; these can be explored in the cellxgene browser (*14*), which

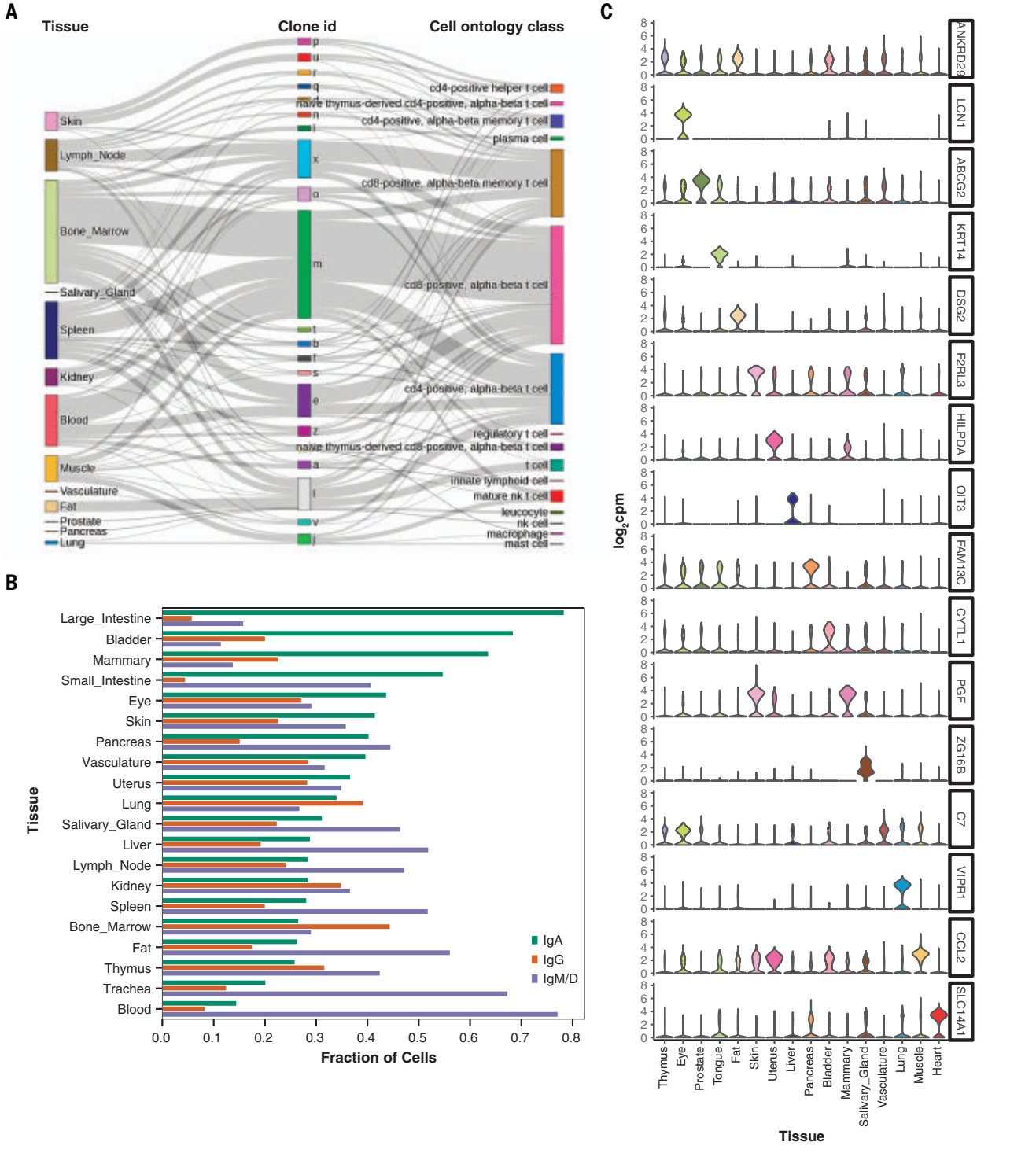


Fig. 3. Analysis of immune and endothelial cell types shared across tissues. (**A**) Illustration of clonal distribution of T cells across multiple tissues. The majority of T cell clones are found in multiple tissues and represent a variety of T cell subtypes. nk cell, natural killer cell. (**B**) Prevalence of B cell isotypes across tissues, ordered by decreasing abundance of IgA. (**C**) Expression levels of tissue-specific endothelial markers, shown as violin plots, in the entire dataset. Many of the markers are highly tissue specific and typically were derived from multiple donors, as follows: bladder (3 donors), eye (2), fat (2), heart (1), liver (2), lung (3), mammary (1), muscle (4), pancreas (2), prostate (2), salivary gland (2), skin (2), thymus (2), tongue (2), uterus (1), and vasculature (2). A detailed donor-tissue breakdown is available in table S2.

uses a statistic called SpliZ (*22*). We focus on two examples of cell type–specific splicing of two well-studied genes: MYL6 and CD47. Similar cell type–specific splice usage was also observed with TPM1, TPM2, and ATP5F1C, three other genes with well-characterized splice variants (fig. S15).

MYL6 is an essential light chain (ELC) for myosin and is highly expressed in all tissues and compartments. Yet, splicing of MYL6, particularly that involving the inclusion or exclusion of exon 6 (Fig. 4A), varies in a cell type– and compartment-specific manner (Fig. 4B). Although the isoform excluding exon 6 has previously been mainly described in phasic

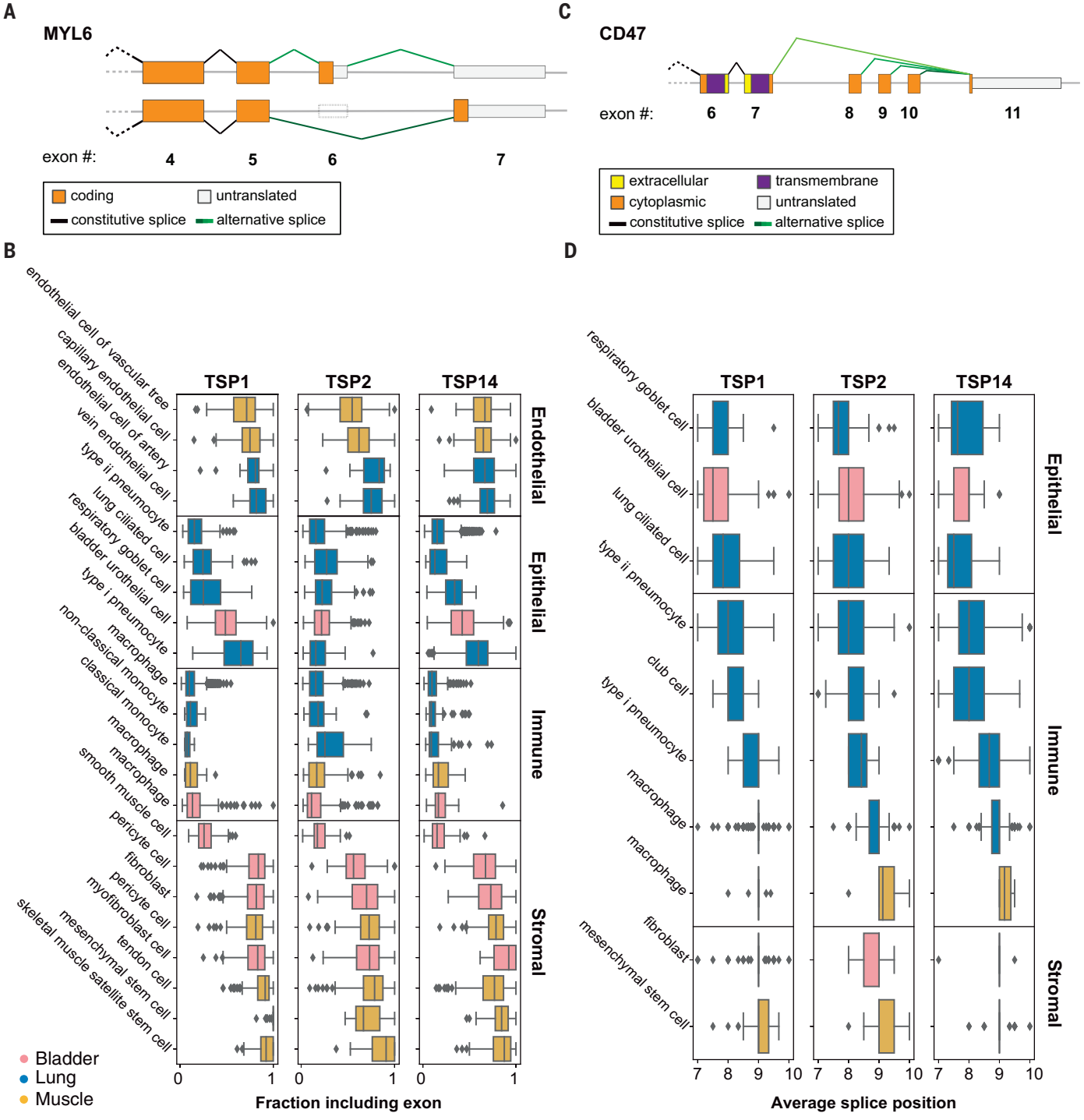


Fig. 4. Alternative splicing analysis. (**A** and **B**) The sixth exon in MYL6 is skipped at different proportions in different compartments. Cells in the immune and epithelial compartments tend to skip the exon, whereas cells in the endothelial and stromal compartments tend to include the exon. Boxes are grouped by compartment and colored by tissue. The fraction of junctional reads that include exon 6 was calculated for each cell with >10 reads mapping to the exon-skipping event. Horizontal box plots in (B) show the distribution of exon inclusion in each cell type. (**C** and **D**) Alternative splicing in CD47 involves one 5′ splice site (exon 11; 108,047,292) and four 3′ splice sites. Horizontal box plots in (D) show the distribution of weighted averages of alternative 3′ splice sites in each cell type. Epithelial cells tend to use exons closer to the 5′ splice site compared with immune and stromal cells. Boxes are grouped by compartment and colored by tissue.

smooth muscle (*23*), we discovered that it can also be the predominant isoform in non–smooth muscle cell types. Our analysis establishes pervasive regulation of MYL6 splicing in many cell types, such as endothelial and immune cells. These previously unknown, compartment-specific expression patterns of the two MYL6 isoforms are reproduced in multiple individuals from the Tabula Sapiens dataset (Fig. 4, A and B).

CD47 is a multispanning membrane protein involved in many cellular processes, including angiogenesis and cell migration and as a "do not eat me" signal to macrophages (*24*). Differential use of exons 7 to 10 (Fig. 4C

and fig. S14F) composes a variably long cytoplasmic tail (*25*). Immune cells—but also stromal and endothelial cells—have a distinct, consistent splicing pattern in CD47 that dominantly excludes two proximal exons and splicing directly to exon 8. In contrast to other compartments, epithelial cells exhibit a different splicing pattern that increases the length of the cytoplasmic tail by splicing more commonly to exon 9 and exon 10 (Fig. 4D). Characterization of the splicing programs of CD47 in single cells may have implications for understanding the differential signaling activities of CD47 and for therapeutic manipulation of CD47 function.

Cell state dynamics can be inferred from a single time point

Although the Tabula Sapiens was created from a single moment in time for each donor, it is possible to infer dynamic information from the data. Cell division is an important transient change of internal cell state, and we computed a cycling index for each cell type to identify actively proliferating versus quiescent or postmitotic cell states. Rapidly dividing progenitor cells had among the highest cycling indices, whereas cell types from the endothelial and stromal compartments, which are known to be largely quiescent, had low cycling indices (Fig. 5A). In intestinal tissue, transient amplifying cells and the crypt stem cells divide rapidly in the intestinal crypts to give rise to terminally differentiated cell types of the villi (*26*). These cells were ranked with the highest cycling indices, whereas terminally differentiated cell types, such as the goblet cells, had the lowest ranks (fig. S16A). To complement the computational analysis of cell cycling, we performed immunostaining of intestinal tissue for the MKI67 protein (commonly referred to as K_i-67) and confirmed that transient amplifying cells abundantly express this proliferation marker (fig. S16, B and C), which supports the conclusion that this marker is differentially expressed in the G2/M cluster.

We observed several interesting tissue-specific differences in cell cycling. To illustrate one example, UMAP clustering of macrophages showed tissue-specific clustering of this cell type and that blood, bone marrow, and lung macrophages have the highest cycling indices compared with macrophages found in the bladder, skin, and muscle (fig. S16, D to G). Consistent with this finding, the expression values of cyclin-dependent kinase (CDK) inhibitors (in particular the gene CDKN1A), which block the cell cycle, have the lowest overall expression in macrophages from tissues with high cycling indices (fig. S16F).

We used RNA velocity (*27*) as a further dynamic approach to study transdifferentiation of bladder mesenchymal cells to myofibroblasts (Fig. 5B). Latent time analysis, which provides an estimate of each cell's internal clock using RNA velocity trajectories (*28*), correctly identified the direction of differentiation (Fig. 5C) across multiple donors. Ordering cells as a function of latent time shows clustering of the mesenchymal and myofibroblast gene expression programs for the most dynamically expressed genes (Fig. 5D). Among these genes, ACTN1 (alpha actinin 1)—a key actin cross-linking protein that stabilizes cytoskeleton-membrane interactions (*29*)—increases across the mesenchymal-to-myofibroblast transdifferentiation trajectory (fig. S16H). Another gene with a similar trajectory is MYLK (myosin light-chain kinase), which also rises as myofibroblasts attain more muscle-like properties (*30*). Finally, a random sampling of the most dynamic genes shared across TSP1 and TSP2 demonstrated that they share concordant trajectories and revealed some of the core genes in the transcriptional program underlying this transdifferentiation event within the bladder (fig. S16I).

Unexpected spatial variation in the microbiome

The Tabula Sapiens provided an opportunity to densely and directly sample the human microbiome throughout the gastrointestinal tract. The intestines from donors TSP2 and TSP14 were sectioned into five regions: the duodenum, jejunum, ileum, and ascending and sigmoid colon (Fig. 6A). Each section was transected, and three to nine samples were collected from each location, followed by amplification and sequencing of the 16*S* ribosomal RNA (rRNA) gene. Uniformly, there was a high (~10 to 30%) relative abundance of Proteobacteria, particularly Enterobacteriaceae (Fig. 6B), even in the colon. Samples from each of the duodenum, jejunum, and ileum were largely distinct from one another, with samples exhibiting individual patterns of blooming or absence of certain families (Fig. 6B). These data reveal that the microbiota are patchy, even at a 3-inch (7.62-cm) length scale. We observed similar heterogeneity in both donors (fig. S17, A to C). In the small intestine, richness (number of observed species) was also variable and was negatively correlated with the relative abundance of Burkholderiaceae (Fig. 6B); in TSP2, the Proteobacteria phylum was dominated by Enterobacteriaceae, which was present at >30% in all samples at a level negatively correlated with richness (fig. S17, A to C). In a comparison of species from adjacent regions across the gut, a large fraction of species was specific to each region (Fig. 6C), reflecting the patchiness. These data are derived from only two donor samples, and further conclusions about the statistics and extent of microbial patchiness will require larger studies.

We analyzed host immune cells in conjunction with the spatial microbiome data; UMAP clustering analysis revealed that the small intestine T cell pool from TSP14 contained a population with distinct transcriptomes (Fig. 6D). The most significant transcriptional differences in T cells between the small and large intestine were genes associated with trafficking, survival, and activation (Fig. 6E and table S8). For example, expression of the long noncoding RNA MALAT1, which affects the regulatory function of T cells, and CCR9, which mediates T lymphocyte development and migration to the intestine (*31*), were high only in the small intestine, whereas GPR15 (colonic T cell trafficking), SELENBP1 (selenium transporter), ANXA1 (repressor of inflammation in T cells), KLRC2 (T cell lectin), CD24 (T cell survival), GDF15 (T cell inhibitor), and RARRES2 (T cell chemokine) exhibited much higher expression in the large intestine. Within the epithelial cells, we observed distinct transcriptomes between small and large intestine Paneth cells and between small and large intestine enterocytes, whereas there was some degree of overlap for each of the two cell types for either location (fig. S17, E and F). The site-specific composition of the microbiome in the intestine, paired with distinct T cell populations at each site, helps define local host-microbe interactions that occur in the gastrointestinal tract and is likely reflective of a gradient of physiological conditions that influence host-microbe dynamics.

Conclusion

The Tabula Sapiens is part of a growing set of data that, when analyzed together, will enable many interesting comparisons of both a biological and technical nature. Studying particular cell types across organs, datasets, and species will yield new biological insights—as shown with fibroblasts (*32*). Similarly, comparing fetal human cell types (*33*) with those determined in this work in adults may give insight into the loss of plasticity from early development to maturity. Having multiorgan data from individual donors may facilitate the development of methods to compare diverse datasets and yield understanding of technical artifacts from various approaches (*8*, *9*, *34*, *35*). The Tabula Sapiens has enabled discoveries relating to shared behavior and organ-specific differences across cell types. For example, we found T cell clones shared between organs and characterized organ-dependent hypermutation rates among resident B cells. Endothelial cells and macrophages are cell types that are shared across tissues but often show subtle tissue-specific differences in gene expression. We found an unexpectedly large and diverse amount of cell type–specific RNA splice variant usage and discovered and validated many previously

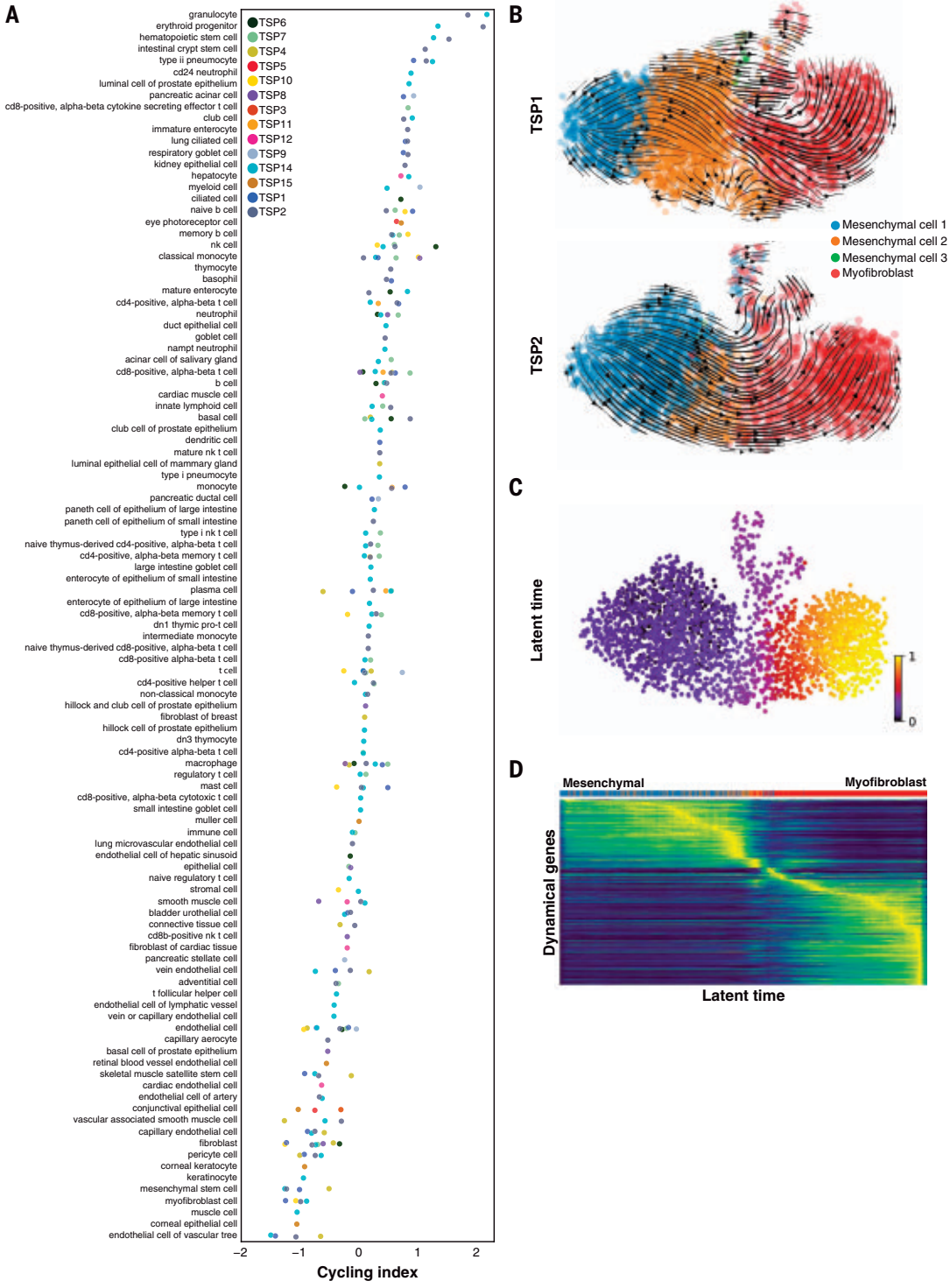


Fig. 5. Dynamic changes in cell state. (**A**) Cell types ordered by magnitude of cell cycling index per donor (each a separate color), with the most highly proliferative at the top and quiescent cells at the bottom of the list. (**B**) RNA velocity analysis demonstrating mesenchymal-to-myofibroblast transition in the bladder. The arrows represent a flow derived from the ratio of unspliced to spliced transcripts, which in turn predicts dynamic changes in cell identity. (**C** and **D**) Latent time analysis of the mesenchymal-to-myofibroblast transition in the bladder, demonstrating stereotyped changes in gene expression trajectory.

undiscovered splices. These are but a few examples of how the Tabula Sapiens represents a broadly useful reference to deeply understand and explore human biology at cellular resolution.

Materials and methods summary

Fresh, whole, and nontransplantable organs, or 1- to 2-cm^3 organ samples, were obtained from surgery and then transported on ice by courier to tissue expert laboratories, where they were immediately prepared for transcriptome sequencing. Single-cell suspensions were prepared for 10× Genomics 3′ V3.1 droplet-based sequencing and for FACS-sorted 384-well plate smart-seq2. Preparation began with

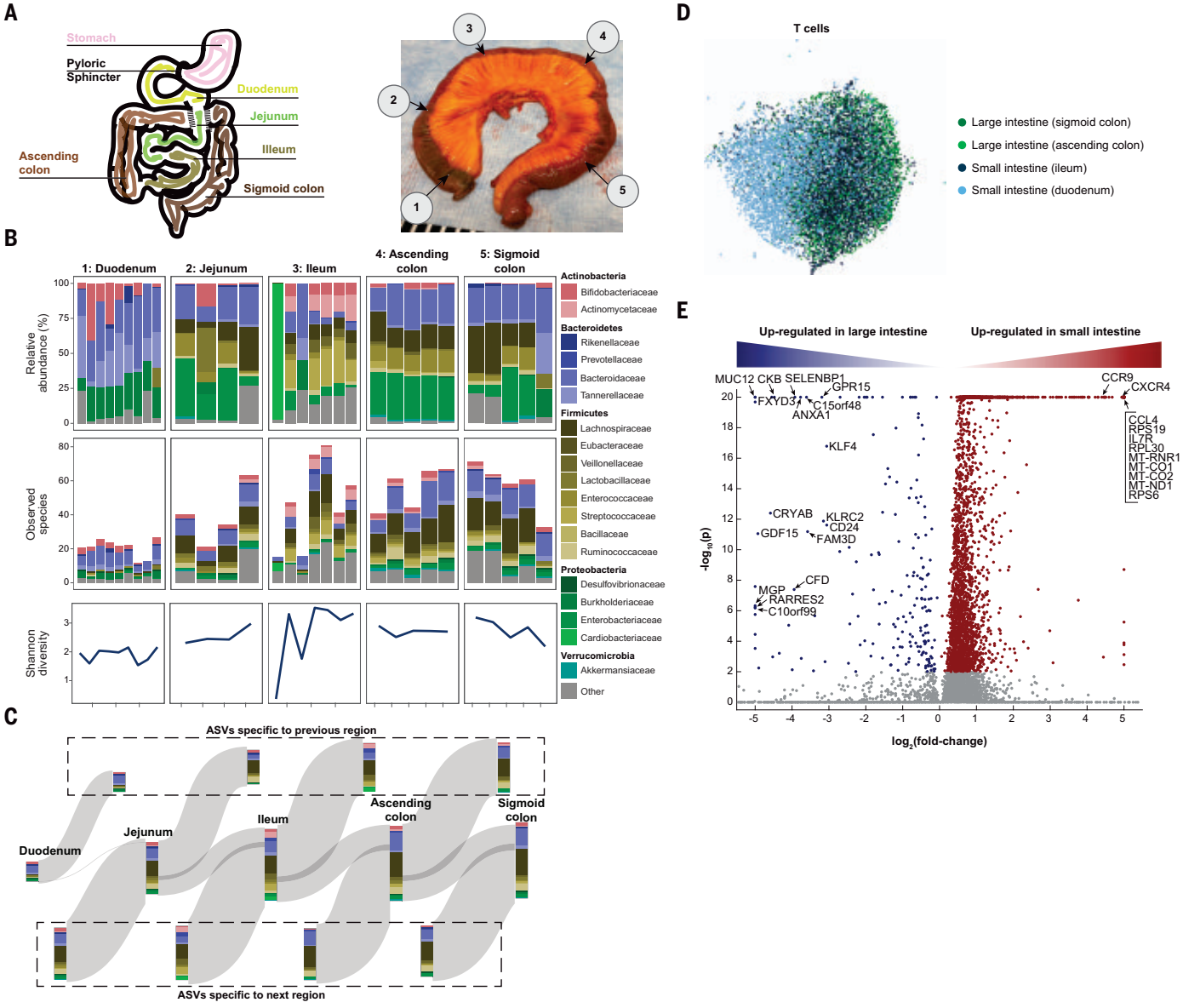


Fig. 6. High-resolution view highlights patchiness of the gut microbiome. (**A**) Schematic (left) and photo (right) of the colon from donor TSP2, with numbers 1 to 5 representing microbiota sampling locations. (**B**) Relative abundances and richness (number of observed species) at the family level in each sampling location, as determined by 16*S* rRNA sequencing. The Shannon diversity, a metric of evenness, mimics richness. Variability in relative abundance and/or richness or Shannon diversity was higher in the duodenum, jejunum, and ileum compared with the ascending and sigmoid colon. (**C**) A Sankey diagram showing the inflow and outflow of microbial species from each section of the gastrointestinal tract. The stacked bar for each gastrointestinal section represents the number of observed species in each family as the union of all sampling locations for that section. The stacked bar flowing out represents gastrointestinal species not found in the subsequent section, and the stacked bar flowing into each gastrointestinal section represents the species not found in the previous section. ASVs, amplicon sequence variants. (**D**) UMAP clustering of T cells reveals distinct transcriptome profiles in the distal and proximal small and large intestines. (**E**) Dots in volcano plot highlight genes up-regulated in the large (left) and small (right) intestines. Labeled dots include genes with known roles in trafficking, survival, and activation.

dissection, digestion with enzymes, and physical manipulation; tissue-specific details are available in the complete materials and methods (*12*). Cell suspensions from some organs were normalized by major cell compartment (epithelial, endothelial, immune, and stromal) using antibody-labeled magnetic microbeads to enrich rare cell types. cDNA and sequencing libraries were prepared and run on the Illumina NovaSeq 6000 with the goal to obtain 10,000 droplet-based cells and 1000 plate-based cells for each organ. Sequences were demultiplexed and aligned to the GRCh38 reference genome. Gene count tables were generated with CellRanger (droplet samples) or STAR and HTSEQ (plate samples). Cells with low unique molecular identifier (UMI) counts or low gene counts were removed. Droplet cells were filtered to remove barcode-hopping events and filtered for ambient RNA using DecontX. Sequencing batches were harmonized using scVI and projected to two-dimensional (2D) space with UMAP for analysis by the tissue experts. Expert annotation was made through the cellxgene browser and regularized with a public cell ontology. Annotation was manually QC checked and cross-validated with PopV, an

annotation tool that uses seven different automated annotation methods. For complete materials and methods, see the supplementary materials (*12*).

REFERENCES AND NOTES

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell* (Garland Science, ed. 4, 2002).
2. A. Regev *et al*., The human cell atlas. *eLife* **6**, e27041 (2017). doi: 10.7554/eLife.27041; pmid: 29206104
3. M. Uhlén *et al*., Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). doi: 10.1126/science.1260419; pmid: 25613900
4. P. J. Thul *et al*., A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017). doi: 10.1126/science.aal3321; pmid: 28495876
5. A. O. Pisco, B. Tojo, A. McGeever, Single-Cell Analysis for Whole-Organism Datasets. *Annu. Rev. Biomed. Data Sci.* **4**, 207–226 (2021). doi: 10.1146/annurev-biodatasci-092820-031008; pmid: 34465173
6. S. Aldridge, S. A. Teichmann, Single cell transcriptomics comes of age. *Nat. Commun.* **11**, 4307 (2020). doi: 10.1038/s41467-020-18158-5; pmid: 32855414
7. A. R. Wu, J. Wang, A. M. Streets, Y. Huang, Single-Cell Transcriptional Analysis. *Annu. Rev. Anal. Chem.* **10**, 439–462 (2017). doi: 10.1146/annurev-anchem-061516-045228; pmid: 28301747
8. X. Han *et al*., Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020). doi: 10.1038/s41586-020-2157-4; pmid: 32214235
9. S. He *et al*., Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biol.* **21**, 294 (2020). doi: 10.1186/s13059-020-02210-0; pmid: 33287869
10. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Principal investigators, Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* **562**, 367–372 (2018). doi: 10.1038/s41586-018-0590-4; pmid: 30283141
11. The Tabula Muris Consortium, A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020). doi: 10.1038/s41586-020-2496-1; pmid: 32669714
12. Materials and methods are available as supplementary materials.
13. S. Wang *et al*., Leveraging the Cell Ontology to classify unseen cell types. *Nat. Commun.* **12**, 5556 (2021). doi: 10.1038/s41467-021-25725-x; pmid: 34548483
14. The Tabula Sapiens Consortium, "Tabula Sapiens Data Portal"; https://tabula-sapiens-portal.ds.czbiohub.org.
15. E. M. Steinert *et al*., Quantifying Memory CD8 T Cells Reveals Regionalization of Immunosurveillance. *Cell* **161**, 737–749 (2015). doi: 10.1016/j.cell.2015.03.031; pmid: 25957682
16. L. C. Davies, P. R. Taylor, Tissue-resident macrophages: Then and now. *Immunology* **144**, 541–548 (2015). doi: 10.1111/imm.12451; pmid: 25684236
17. D. I. Godfrey, H.-F. Koay, J. McCluskey, N. A. Gherardin, The biology and functional importance of MAIT cells. *Nat. Immunol.* **20**, 1110–1128 (2019). doi: 10.1038/s41590-019-0444-8; pmid: 31406380
18. M. Karlsson *et al*., A single-cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, eabh2169 (2021). doi: 10.1126/sciadv.abh2169; pmid: 34321199
19. A. Gillich *et al*., Capillary cell-type specialization in the alveolus. *Nature* **586**, 785–789 (2020). doi: 10.1038/s41586-020-2822-7; pmid: 33057196
20. R. Dehghannasiri, J. E. Olivieri, A. Damljanovic, J. Salzman, Specific splice junction detection in single cells with SICILIAN. *Genome Biol.* **22**, 219 (2021). doi: 10.1186/s13059-021-02434-8; pmid: 34353340
21. J. Lonsdale *et al*., The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013). doi: 10.1038/ng.2653; pmid: 23715323
22. J. E. Olivieri *et al*., RNA splicing programs define tissue compartments and cell types at single-cell resolution. *eLife* **10**, e70692 (2021). doi: 10.7554/eLife.70692; pmid: 34515025
23. S. A. Fisher, Vascular smooth muscle phenotypic diversity and function. *Physiol. Genomics* **42A**, 169–187 (2010). doi: 10.1152/physiolgenomics.00111.2010; pmid: 20736412
24. A. A. Barkal *et al*., CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. *Nature* **572**, 392–396 (2019). doi: 10.1038/s41586-019-1456-0; pmid: 31367043
25. S. M. G. Hayat *et al*., CD47: Role in the immune system and application to cancer therapy. *Cell. Oncol.* **43**, 19–30 (2020). doi: 10.1007/s13402-019-00469-5; pmid: 31485984
26. H. Gehart, H. Clevers, Tales from the crypt: New insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019). doi: 10.1038/s41575-018-0081-y; pmid: 30429586
27. G. La Manno *et al*., RNA velocity of single cells. *Nature* **560**, 494–498 (2018). doi: 10.1038/s41586-018-0414-6; pmid: 30089906
28. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020). doi: 10.1038/s41587-020-0591-3; pmid: 32747759
29. A. C. H. Murphy, P. W. Young, The actinin family of actin cross-linking proteins – A genetic perspective. *Cell Biosci.* **5**, 49 (2015). doi: 10.1186/s13578-015-0029-7; pmid: 26312134
30. R. Li, X. Li, J. Hagood, M.-S. Zhu, X. Sun, Myofibroblast contraction is essential for generating and regenerating the gas-exchange surface. *J. Clin. Invest.* **130**, 2859–2871 (2020). doi: 10.1172/JCI132189; pmid: 32338642
31. S. Uehara, A. Grinberg, J. M. Farber, P. E. Love, A role for CCR9 in T lymphocyte development and migration. *J. Immunol.* **168**, 2811–2819 (2002). doi: 10.4049/jimmunol.168.6.2811; pmid: 11884450
32. M. B. Buechler *et al*., Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021). doi: 10.1038/s41586-021-03549-5; pmid: 33981032
33. J. Cao *et al*., A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020). doi: 10.1126/science.aba7721; pmid: 33184181
34. G. Eraslan *et al*., Single-nucleus cross-tissue molecular reference maps to decipher disease gene function. *bioRxiv*2021.07.19.452954 [Preprint] (2021). doi: 10.1101/2021.07.19.452954
35. C. Domínguez Conde *et al*., Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture in humans. *bioRxiv*2021.04.28.441762 [Preprint] (2021). doi: 10.1101/2021.04.28.441762
36. The Tabula Sapiens Consortium, czbiohub/tabula-sapiens: v1.0, version manuscript1, Zenodo (2022); doi: 10.5281/zenodo.6069683
37. The Tabula Sapiens Consortium, Tabula Sapiens Single-Cell Dataset, version 4, figshare (2021); doi: 10.6084/m9.figshare.14267219.v4
38. The Tabula Sapiens Consortium, Tabula Sapiens H&E Image Collection, version 2, figshare (2021); doi: 10.6084/m9.figshare.14962947.v2
39. The Tabula Sapiens Consortium, Tabula Sapiens Splicing, *figshare* (2021); doi: 10.6084/m9.figshare.14977281.v1

ACKNOWLEDGMENTS

We express our gratitude and thanks to donor WEM and his family, as well as to all the anonymous organ and tissue donors and their families for giving both the gift of life and the gift of knowledge through their generous donations. We also thank Donor Network West for their partnership on this project, the UCSF Liver Center (funded by NIH P30DK026743) for assistance with the liver cell isolations, S. Schmid for a close reading of the manuscript, and B. Tojo for the original artwork in Fig. 1. The data portal for this publication is available at (*14*). **Funding:** This project has been made possible in part by grant nos. 2019-203354, 2020-224249, and 2021-237288 from the Chan Zuckerberg Initiative DAF; an advised fund of Silicon Valley Community Foundation; and by support from the Chan Zuckerberg Biohub. **Author contributions:** See below, where authors are arranged by working group and contributions. **Competing interests:** N. Yosef is an advisor and/or has equity in for Cellarity, Celsius Therapeutics, and Rheos Medicines. The authors declare no other competing interests. **Data and materials availability:** The entire dataset can be explored interactively at https://tabula-sapiens-portal.ds.czbiohub.org/ (*14*). The code used for the analysis is available from Zenodo (*36*). Gene counts and metadata are available from figshare (*37*) and have been deposited in the Gene Expression Omnibus (GSE201333); the raw data files are available from a public AWS S3 bucket (https://registry.opendata.aws/tabula-sapiens/), and instructions on how to access the data have been provided in the project GitHub. The histology images are available from figshare (*38*). SpliZ scores are available from figshare (*39*). To preserve the donors' genetic privacy, we require a data transfer agreement to receive the raw sequence reads. The data transfer agreement is available upon request.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4896
Materials and Methods
Figs. S1 to S17
Tables S1 to S9
References (*40*–*94*)
MDAR Reproducibility Checklist

The Tabula Sapiens Consortium Overall project direction and coordination: Robert C. Jones[1], Jim Karkanias[2], Mark A. Krasnow[3,4], Angela Oliveira Pisco[2], Stephen R. Quake[1,2,5], Julia Salzman[3,6], Nir Yosef[2,7,8,9]
Donor recruitment: Bryan Bulthaup[10], Phillip Brown[10], William Harper[10], Marisa Hemenez[10], Ravikumar Ponnusamy[10], Ahmad Salehi[10], Bhavani A. Sanagavarapu[10], Eileen Spallino[10]
Surgeons: Ksenia A. Aaron[11], Waldo Concepcion[10], James M. Gardner[12,13], Burnett Kelly[10,14], Nikole Neidlinger[10], Zifa Wang[10]
Logistical coordination: Sheela Crasta[1,2], Saroja Kolluru[1,2], Maurizio Morri[2], Angela Oliveira Pisco[2], Serena Y. Tan[15], Kyle J. Travaglini[3], Chenling Xu[7]
Organ processing: Marcela Alcántara-Hernández[16], Nicole Almanzar[17], Jane Antony[18], Benjamin Beyersdorf[19], Deviana Burhan[20], Kruti Calcuttawala[21], Matthew M. Carter[16], Charles K. F. Chan[18,22], Charles A. Chang[23], Stephen Chang[3,19], Alex Colville[21,24], Sheela Crasta[1,2], Rebecca N. Culver[25], Ivana Cvijović[1,5], Gaetano D'Amato[26], Camille Ezran[3], Francisco X. Galdos[18], Astrid Gillich[3], William R. Goodyer[27], Yan Hang[23,28], Alyssa Hayashi[1], Sahar Houshdaran[29], Xianxi Huang[19,30], Juan C. Irwin[29], SoRi Jang[3], Julia Vallve Juanico[29], Aaron M. Kershner[18], Soochi Kim[21,24], Bernhard Kiss[18], Saroja Kolluru[1,2], William Kong[18], Maya E. Kumar[17], Angera H. Kuo[18], Rebecca Leylek[16], Baoxiang Li[31], Gabriel B. Loeb[32], Wan-Jin Lu[18], Sruthi Mantri[33], Maxim Markovic[1], Patrick L. McAlpine[11,34], Antoine de Morree[21,24], Maurizio Morri[2], Karim Mrouj[18], Shravani Mukherjee[31], Tyler Muser[17], Patrick Neuhöfer[3,35,36], Thi D. Nguyen[37], Kimberly Perez[16], Ragini Phansalkar[26], Angela Oliveira Pisco[2], Nazan Puluca[18], Zhen Qi[18], Poorvi Rao[20], Hayley Raquer-McKay[16], Nicholas Schaum[18,21], Bronwyn Scott[31], Bobak Seddighzadeh[38], Joe Segal[20], Sushmita Sen[29], Shaheen Sikandar[18], Sean P. Spencer[16], Lea C. Steffes[17], Varun R. Subramaniam[31], Aditi Swarup[31], Michael Swift[1], Kyle J. Travaglini[3], Will Van Treuren[16], Emily Trimm[26], Stefan Veizades[19,39], Sivakamasundari Vijayakumar[18], Kim Chi Vo[29], Sevahn K. Vorperian[1,40], Wanxin Wang[29], Hannah N. W. Weinstein[38], Juliane Winkler[41], Timothy T. H. Wu[3], Jamie Xie[38], Andrea R. Yung[3], Yue Zhang[3]
Sequencing: Angela M. Detweiler[2], Honey Mekonen[2], Norma F. Neff[2], Rene V. Sit[2], Michelle Tan[2], Jia Yan[2]
Histology: Gregory R. Bean[15], Vivek Charu[15], Erna Forgó[15], Brock A. Martin[15], Michael G. Ozawa[15], Oscar Silva[15], Serena Y. Tan[15], Angus Toland[15], Venkata N. P. Vemuri[2]
Data analysis: Shaked Afik[7], Kyle Awayan[2], Olga Borisovna Botvinnik[2], Ashley Byrne[2], Michelle Chen[1], Roozbeh Dehghannasiri[3,6], Angela M. Detweiler[2], Adam Gayoso[7], Alejandro A. Granados[2], Qiqing Li[2], Gita Mahmoudabadi[1], Aaron McGeever[2], Antoine de Morree[21,24], Julia Eve Olivieri[3,6,42], Madeline Park[2], Angela Oliveira Pisco[2], Neha Ravikumar[1], Julia Salzman[3,6], Geoff Stanley[1], Michael Swift[1], Michelle Tan[2], Weilun Tan[2], Alexander J. Tarashansky[2], Rohan Vanheusden[2], Sevahn K. Vorperian[1,40], Peter Wang[3,6], Sheng Wang[2], Galen Xing[2], Chenling Xu[6], Nir Yosef[2,6,7,8]
Expert cell type annotation: Marcela Alcántara-Hernández[16], Jane Antony[18], Charles K. F. Chan[18,22], Charles A. Chang[23], Alex Colville[21,24], Sheela Crasta[1,2], Rebecca Culver[25], Les Dethlefsen[43], Camille Ezran[3], Astrid Gillich[3], Yan Hang[23,28], Po-Yi Ho[16], Juan C. Irwin[29], SoRi Jang[3], Aaron M. Kershner[18], William Kong[18], Maya E. Kumar[17], Angera H. Kuo[18], Rebecca Leylek[16], Shixuan Liu[3,44], Gabriel B. Loeb[32], Wan-Jin Lu[18], Jonathan S. Maltzman[45,46], Ross J. Metzger[27,47], Antoine de Morree[21,24], Patrick Neuhöfer[3,35,36], Kimberly Perez[16], Ragini Phansalkar[26], Zhen Qi[18], Poorvi Rao[20], Hayley Raquer-McKay[16], Koki Sasagawa[19], Bronwyn Scott[31], Rahul Sinha[15,18,35], Hanbing Song[38], Sean P. Spencer[16], Aditi Swarup[31], Michael Swift[1], Kyle J. Travaglini[3], Emily Trimm[26], Stefan Veizades[19,39], Sivakamasundari Vijayakumar[18], Bruce Wang[20], Wanxin Wang[29], Juliane Winkler[41], Jamie Xie[38], Andrea R. Yung[3]
Tissue expert principal investigators: Steven E. Artandi[3,35,36], Philip A. Beachy[18,23,48], Michael F. Clarke[18], Linda C. Giudice[29], Franklin W. Huang[38,49], Kerwyn Casey Huang[1,16], Juliana Idoyaga[16], Seung K. Kim[23,28], Mark Krasnow[3,4], Christin S. Kuo[17],

Patricia Nguyen[19,39,46], Stephen R. Quake[1,2,5], Thomas A. Rando[21,24], Kristy Red-Horse[26], Jeremy Reiter[50], David A. Relman[16,43,46], Justin L. Sonnenburg[16], Bruce Wang[20], Albert Wu[31], Sean M. Wu[19,39], Tony Wyss-Coray[21,24]

[1]Department of Bioengineering, Stanford University, Stanford, CA, USA. [2]Chan Zuckerberg Biohub, San Francisco, CA, USA. [3]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. [4]Howard Hughes Medical Institute, USA. [5]Department of Applied Physics, Stanford University, Stanford, CA, USA. [6]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. [7]Center for Computational Biology, University of California Berkeley, Berkeley, CA, USA. [8]Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA, USA. [9]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [10]Donor Network West, San Ramon, CA, USA. [11]Department of Oto-laryngology-Head and Neck Surgery, Stanford University School of Medicine, Stanford, California, USA. [12]Department of Surgery, University of California San Francisco, San Francisco, CA, USA. [13]Diabetes Center, University of California San Francisco, San Francisco, CA, USA. [14]DCI Donor Services, Sacramento, CA, USA. [15]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. [16]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. [17]Department of Pediatrics, Division of Pulmonary Medicine, Stanford University, Stanford, CA, USA. [18]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA. [19]Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA. [20]Department of Medicine and Liver Center, University of California San Francisco, San Francisco, CA, USA. [21]Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA. [22]Department of Surgery – Plastic and Reconstructive Surgery, Stanford University School of Medicine, Stanford, CA, USA. [23]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA. [24]Paul F. Glenn Center for the Biology of Aging, Stanford University School of Medicine, Stanford, CA, USA. [25]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [26]Department of Biology, Stanford University, Stanford, CA, USA. [27]Department of Pediatrics, Division of Cardiology, Stanford University School of Medicine, Stanford, CA, USA. [28]Stanford Diabetes Research Center, Stanford University School of Medicine, Stanford, CA, USA. [29]Center for Gynecology and Reproductive Sciences, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, San Francisco, CA, USA. [30]Department of Critical Care Medicine, The First Affiliated Hospital of Shantou University Medical College, Shantou, China. [31]Department of Ophthalmology, Stanford University School of Medicine, Stanford, CA, USA. [32]Division of Nephrology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA. [33]Stanford University School of Medicine, Stanford, CA, USA. [34]Mass Spectrometry Platform, Chan Zuckerberg Biohub, Stanford, CA, USA. [35]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. [36]Department of Medicine, Division of Hematology, Stanford University School of Medicine, Stanford, CA, USA. [37]Department of Biochemistry and Biophysics, Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, USA. [38]Division of Hematology and Oncology, Department of Medicine, Bakar Computational Health Sciences Institute, Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA. [39]Stanford Cardiovascular Institute, Stanford CA, USA. [40]Department of Chemical Engineering, Stanford University, Stanford, CA, USA. [41]Department of Cell & Tissue Biology, University of California San Francisco, San Francisco, CA, USA. [42]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA. [43]Division of Infectious Diseases & Geographic Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [44]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA. [45]Division of Nephrology, Stanford University School of Medicine, Stanford, CA, USA. [46]Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA. [47]Vera Moulton Wall Center for Pulmonary and Vascular Disease, Stanford University School of Medicine, Stanford, CA, USA. [48]Department of Urology, Stanford University School of Medicine, Stanford, CA, USA. [49]Division of Hematology/Oncology, Department of Medicine, San Francisco Veterans Affairs Health Care System, San Francisco, CA, USA. [50]Department of Biochemistry, University of California San Francisco, San Francisco, CA, USA.

Submitted 19 July 2021; accepted 31 March 2022
10.1126/science.abl4896

RESEARCH ARTICLE SUMMARY

HUMAN CELL ATLAS

Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function

Gökcen Eraslan†, Eugene Drokhlyansky†, Shankara Anand‡, Evgenij Fiskin‡, Ayshwarya Subramanian‡, Michal Slyper‡, Jiali Wang‡, Nicholas Van Wittenberghe, John M. Rouhana, Julia Waldman, Orr Ashenberg, Monkol Lek, Danielle Dionne, Thet Su Win, Michael S. Cuoco, Olena Kuksenko, Alexander M. Tsankov, Philip A. Branton, Jamie L. Marshall, Anna Greka, Gad Getz, Ayellet V. Segrè*§, François Aguet*§, Orit Rozenblatt-Rosen*§, Kristin G. Ardlie*§, Aviv Regev*§

INTRODUCTION: Understanding and treating disease requires deep, systematic characterization of different cells and their interactions across human tissues and organs, along with characterization of the genetic variants that causally contribute to disease risk. Recent studies have combined single-cell atlases of specific human tissues and organs with genes associated with human disease to relate risk variants to likely cells of action. However, it has been challenging to extend these studies to profile multiple tissues and organs across the body, conduct studies at population scale, and integrate cell atlases from multiple organs to yield unified insights.

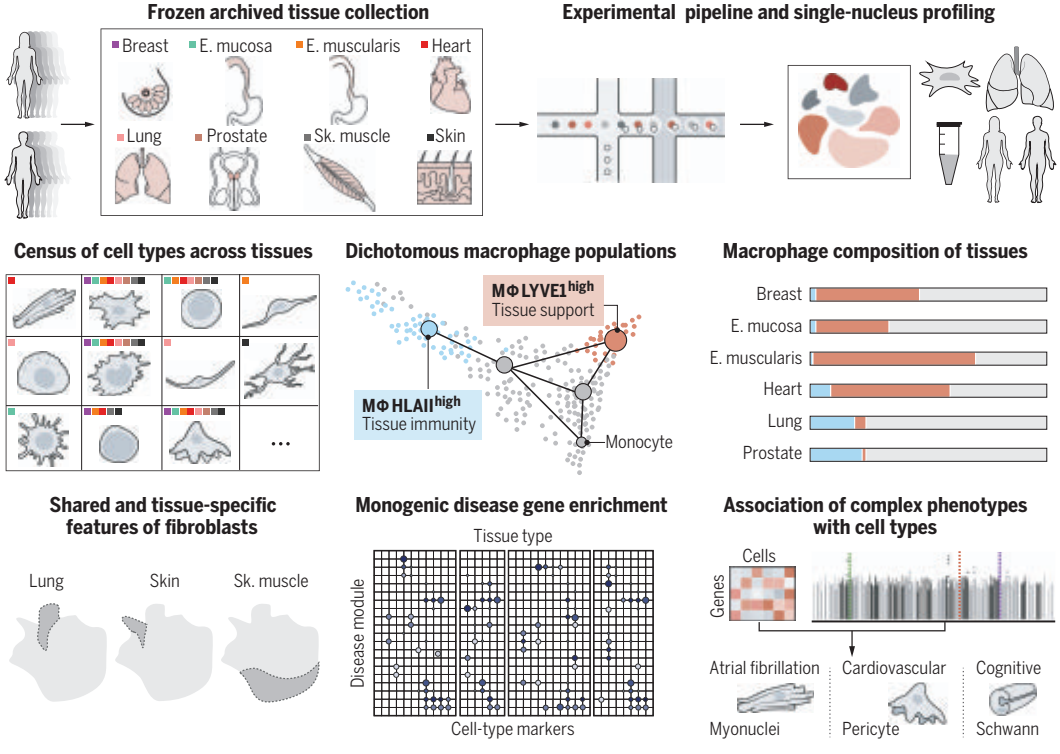
RATIONALE: Because of the pleiotropy and specificity of disease-associated variants, systematically relating variants to cells and molecular processes requires analysis across multiple tissues and individuals. Prior cell atlases primarily relied on fresh tissue samples from a single organ or tissue. Single-nucleus RNA sequencing (snRNA-seq) can be applied to frozen, archived tissue and captures cell types that do not survive dissociation across many tissues. Deep learning methods can integrate data across individuals and tissues by controlling for batch effects while preserving biological variation.

RESULTS: We established a framework for multitissue human cell atlases and generated an atlas of 209,126 snRNA-seq profiles from eight tissue types across 16 individuals, archived as frozen tissue as part of the Genotype-Tissue Expression (GTEx) project. We benchmarked four protocols and show how to apply them in a pooled setting to enable larger studies. We integrated the cross-tissue atlas using a conditional variational autoencoder, annotated it with 43 broad and 74 fine categories, and demonstrated its use to decipher tissue residency, such as a macrophage dichotomy and lipid associations that are preserved across tissues, and tissue-specific fibroblast features, including lung alveolar fibroblasts with likely roles in mechanosensation. We relate cells to human disease biology and disease-risk genes for both rare and common diseases, including rare muscle disease gene groups enriched in distinct subsets of myonuclei and nonmyocytes, and cell type–specific enrichment of expression and splicing quantitative trait locus (QTL) target genes mapped to genome-wide association study loci.

CONCLUSION: Our framework will empower large, cross-tissue population and/or disease studies at single-cell resolution. These frameworks and the cross-tissue perspective provided here will form a basis for larger-scale future studies to improve our understanding of cross-tissue and cross-individual variation of cellular phenotypes in relation to disease-associated genetic variation. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ayellet_segre@meei.harvard.edu (A.V.S.); francois@broadinstitute.org (F.A.); orit.r.rosen@gmail.com (O.R.-R.); kardlie@broadinstitute.org (K.G.A.); aviv.regev.sc@gmail.com (A.R.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
§These authors contributed equally to this work.
Cite this article as G. Eraslan *et al.*, *Science* **376**, eabl4290 (2022). DOI: 10.1126/science.abl4290

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl4290



Cross-tissue snRNA-seq atlas in eight frozen, archived adult human tissues. Tissue sites and experimental pipeline (top row). The resulting atlas enables a cross-tissue census of tissue-specific and shared cell types (middle left). Differentiation trajectories and compositional analysis of dichotomous macrophage populations improve our understanding of tissue residency (middle center and right). Analyses of fibroblasts across tissues reveal tissue-specific and shared fibroblast features and their functional interpretation (bottom left). Robust and scalable computational methods enable comprehensive associations of monogenic and complex diseases to tissue-specific and shared cell populations (bottom center and right). E. mucosa, esophagus mucosa; E. muscularis, esophagus muscularis; Sk. muscle, skeletal muscle.

RESEARCH ARTICLE

HUMAN CELL ATLAS

Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function

Gökcen Eraslan[1]†‡, Eugene Drokhlyansky[1]†, Shankara Anand[2]§, Evgenij Fiskin[1]§, Ayshwarya Subramanian[1]§, Michal Slyper[1]§, Jiali Wang[3,4,5]§, Nicholas Van Wittenberghe[1], John M. Rouhana[3,4,5], Julia Waldman[1], Orr Ashenberg[1], Monkol Lek[6], Danielle Dionne[1], Thet Su Win[7], Michael S. Cuoco[1], Olena Kuksenko[1], Alexander M. Tsankov[8], Philip A. Branton[9], Jamie L. Marshall[2], Anna Greka[2,10], Gad Getz[2,11,12], Ayellet V. Segrè[3,4,5]¶*, François Aguet[2]¶#*, Orit Rozenblatt-Rosen[1]‡¶*, Kristin G. Ardlie[2]¶*, Aviv Regev[1,13]‡¶*

Understanding gene function and regulation in homeostasis and disease requires knowledge of the cellular and tissue contexts in which genes are expressed. Here, we applied four single-nucleus RNA sequencing methods to eight diverse, archived, frozen tissue types from 16 donors and 25 samples, generating a cross-tissue atlas of 209,126 nuclei profiles, which we integrated across tissues, donors, and laboratory methods with a conditional variational autoencoder. Using the resulting cross-tissue atlas, we highlight shared and tissue-specific features of tissue-resident cell populations; identify cell types that might contribute to neuromuscular, metabolic, and immune components of monogenic diseases and the biological processes involved in their pathology; and determine cell types and gene modules that might underlie disease mechanisms for complex traits analyzed by genome-wide association studies.

Tissue homeostasis and pathology arise from an intricate interplay between different cell types, such that disease risk is influenced by variation in genes that affect the cells' functions and interactions. Human genetics studies, to date, have mapped tens of thousands of loci that either underlie rare monogenic disease or are associated with complex polygenic disease risk (*1*–*3*), including many in regulatory regions, whereas single-cell genomics has become instrumental in constructing cell atlases of both healthy organs and diseased tissues (*4*–*6*).

Coupling these advances in human genetics and single-cell genomics should enhance our understanding of cell type–specific changes in the function and regulation of disease genes. In particular, tissue (*7*), cell type (*8*–*11*), time point, and stimulation (*12*–*14*) all affect gene expression in disease-associated genetic loci. Recently, studies combining single-cell expression atlases and genetic signals have been able to associate risk genes with specific cell types and states in relevant tissues (*15*–*18*).

Because complex diseases often manifest in and implicate cells across multiple tissues, fully understanding the way in which genetic variation affects disease requires generating atlases from diverse tissues across the body and from many individuals, spanning different populations. This poses several challenges. First, collecting fresh tissue samples at scale is logistically challenging, and some tissues, such as brain, muscle, and adipose, are difficult to process into single-cell suspensions (*19*–*23*). As a result, large-scale single-cell profiling studies in human populations (*24*, *25*) have focused on peripheral blood mononuclear cells, which can be frozen and thawed for multiplexed single-cell analysis. Single-nucleus RNA sequencing (snRNA-seq) offers a compelling alternative because it can be applied to archived, frozen tissues (*26*, *27*) from multiple organs and captures diverse cell types. Second, annotation and classification of cell types and states require defining biological relationships between parenchymal, immune, and stromal cells across tissue types. Finally, data integration and interpretation require cross-tissue analytical frameworks to remove unwanted variation while preserving biological differences; identify cell types and states; and relate cell types and states to monogenic and complex trait genetics.

Results

A multitissue, multi-individual single-nucleus reference atlas from archived, frozen human tissues

We constructed a cross-tissue snRNA-seq atlas from 25 archived, frozen tissue samples, previously collected and banked by the Genotype-Tissue Expression (GTEx) project (*7*), that span three or four samples from each of eight tissue sites—breast, esophagus mucosa, esophagus muscularis, heart, lung, prostate, skeletal muscle, and skin—from 16 individuals (seven males and nine females) (Fig. 1A). We selected the samples by RNA quality, tissue autolysis score, and the availability of existing bulk RNA-seq and genome sequencing data [(*28*); table S1]. Histology slides corresponding to each tissue were reviewed by a pathologist to provide detailed annotations (table S1). Because different nucleus extraction protocols can be optimal for different tissues (*26*, *29*), we isolated nuclei from each sample using four protocols that vary in detergents, salt, buffer, and mechanical preparation methodology [CST (0.49% CHAPS detergent, salts, and Tris buffer), NST (NP-40, salts, and Tris buffer), TST (0.03% Tween 20 detergent, salts, and Tris buffer), and the EZ nuclei isolation kit (proprietary composition; NUC101, Sigma-Aldrich); (*26*, *28*, *29*); table S1], followed by droplet-based single-cell RNA-seq (scRNA-seq) (*28*).

We processed the initial snRNA-seq profiles to retain high-quality nuclei profiles and remove the effects of contaminant transcripts from ambient RNA (*28*). In breast and skin, the majority of nuclei profiles were recovered from only one individual sample for each tissue (breast: 61.3%, skin: 93.1%; table S1). Some tissues and protocols had higher ambient RNA contamination, reflected as spurious expression of highly expressed transcripts from one cell type in nuclei profiles of other cell types. Such effects were more prominent in skeletal muscle and heart [false discovery rate (FDR) < 0.05], irrespective of protocol, but were also present in other tissues [(*28*); fig. S1]. We corrected for ambient RNA contamination with CellBender v2.1 (*30*) (fig. S1) and further applied standard quality control metrics (*28*), retaining 209,126 nuclei profiles across the eight tissues, with a mean of 918 genes and 1519 transcripts (unique molecular identifiers) detected per profile.

Cross-tissue atlas annotation recovers diverse cell types, including difficult-to-profile and rare cell subsets

We integrated data from all samples and methods using a conditional variational autoencoder

[1]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [2]The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [3]Department of Ophthalmology, Harvard Medical School, Boston, MA 02115, USA. [4]Ocular Genomics Institute, Department of Ophthalmology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA 02114, USA. [5]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [6]Department of Genetics, Yale School of Medicine, New Haven, CT 06510, USA. [7]Department of Dermatology, Brigham and Women's Hospital, Boston, MA 02115, USA. [8]Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [9]The Joint Pathology Center Gynecologic/Breast Pathology, Silver Spring, MD 20910, USA. [10]Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA. [11]Center for Cancer Research and Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA. [12]Harvard Medical School, Boston, MA 02115, USA. [13]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

***Corresponding author. Email: ayellet_segre@meei.harvard.edu (A.V.S.); francois@broadinstitute.org (F.A.); orit.r.rosen@gmail.com (O.R.-R.); kardlie@broadinstitute.org (K.G.A.); aviv.regev.sc@gmail.com (A.R.)**

†These authors contributed equally to this work.

‡Present address: Genentech, South San Francisco, CA 94080, USA.

§These authors contributed equally to this work.

¶These authors contributed equally to this work.

#Present address: Illumina Artificial Intelligence Laboratory, Illumina, Inc., San Diego, CA 92121, USA.

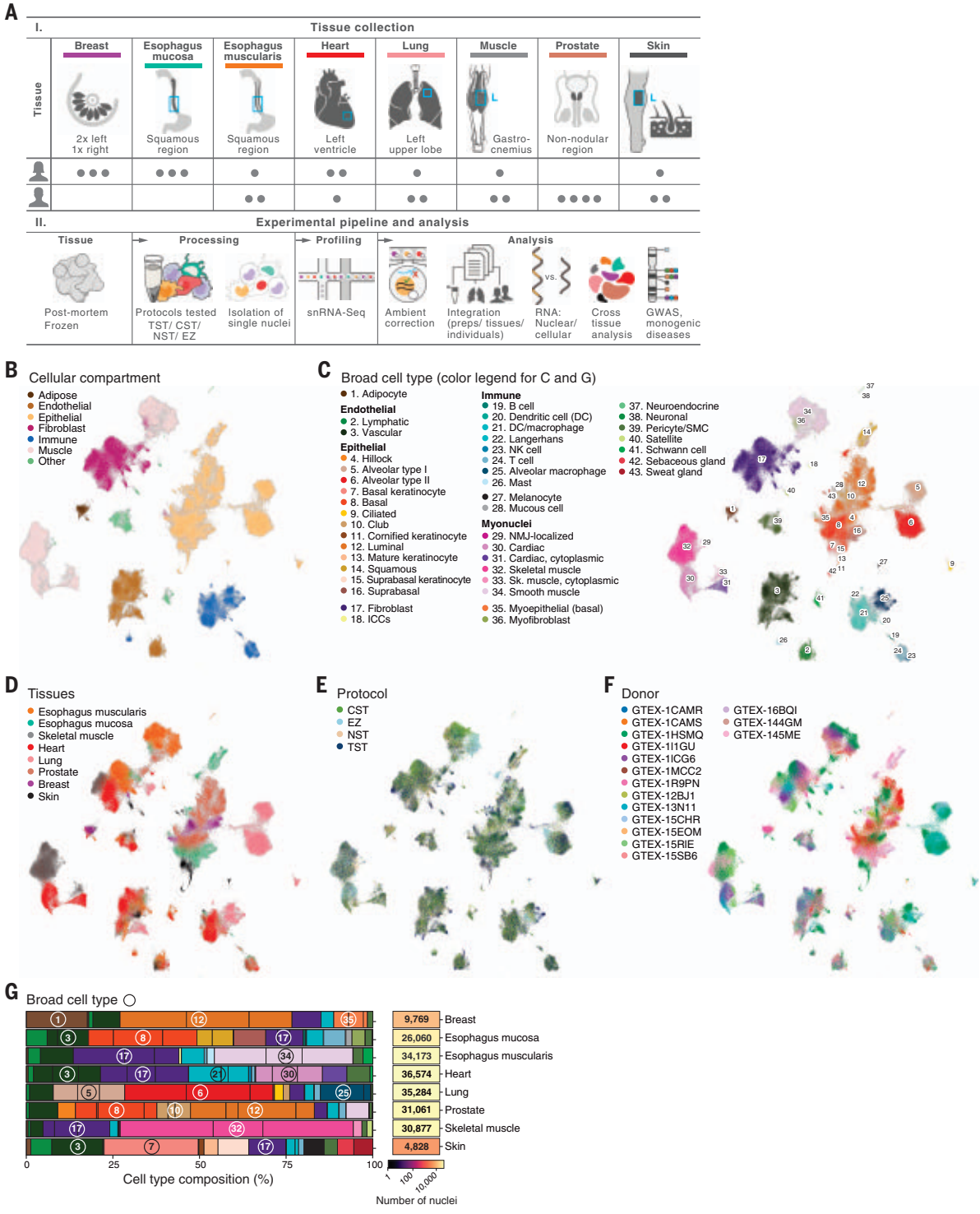


Fig. 1. Cross-tissue snRNA-seq atlas in eight archived, frozen adult human tissues. (**A**) Study design. (**B** to **F**) Cross-tissue single-nucleus atlas. Uniform manifold approximation and projection (UMAP) representation of single-nucleus profiles (dots) colored by main compartments (B), broad cell types (C), tissues (D), isolation protocol (E), and individual donors (F). (**G**) Cell-type composition across tissues. The overall proportion of cells (%) of each type and number of nuclei profiled in each tissue (rows) are shown. Numbers in circles indicate the corresponding broad cell type. Black vertical lines indicate the relative proportion of nuclei from each individual.

(cVAE), which is designed to correct for multiple sources of variation in expression, such as individual-, sex-, and protocol-specific effects, while preserving tissue- and cell type–specific variation [(*28*); Fig. 1, B to F, and figs. S2A and S3]. We benchmarked the cVAE against several other data integration methods, obtaining comparable or improved results and providing guidelines for future integration efforts [fig. S4 and supplementary text note S1; (*28*)]. Cells grouped first by cell type and then by tissue-specific subclusters (Fig. 1, B to D), suggesting that the variation between cell types is larger than the variation within a cell type across tissues.

We annotated cell types within each tissue after dimensionality reduction and graph-based clustering (*28*) by identifying genes that are differentially expressed between clusters and comparing them with literature-based marker genes [(*28*); tables S2 and S3]. We curated comprehensive lists of cell-type markers from the literature for each tissue (figs. S5 and S6 and table S3), including markers for relatively poorly characterized cells, such as interstitial cells of Cajal (ICCs). We defined cellular compartments shared across tissues (e.g., adipose, endothelial, epithelial, fibroblast, immune, muscle) (Fig. 1B), broad cell types (e.g., luminal epithelial cells, vascular endothelial cells) (Fig. 1C and fig. S5), and granular cell subsets (e.g., luminal epithelial cell 1 and 2) (fig. S6). The annotations were consistent across extraction protocols, tissues, and donors (figs. S2, B and D, and S7).

The atlas features 43 broad cell classes (Fig. 1C and tables S2 and S3), with both tissue-shared cell types and tissue-specific subsets (e.g., Fig. 1G and figs. S2, B and D, and S5). For example, tissue-specific cell types such as pneumocytes (alveolar type I and II) and keratinocytes were the predominant cell types in the lung and skin, respectively. Many shared broad cell types such as immune and stromal cells were detected across all tissues (fig. S2, D and E), but with tissue-specific specializations (discussed later in the text). For example, macrophages made up the largest immune population, with diverse subsets of tissue-resident cells.

The atlas captured profiles from cell classes that are difficult to profile by dissociation-based scRNA-seq (*23*, *31*, *32*), including 2350 adipocytes, 21,607 skeletal muscle myonuclei, and 9619 cardiac myonuclei. We detected adipocytes in five of the eight tissue types (breast, muscle, heart, esophagus muscularis, and skin), with 86% of adipocytes from breast tissue (fig. S2D), making up 18% of all breast nuclei profiles (Fig. 1G) (*28*). Skeletal and cardiac myonuclei included key subsets (*33*, *34*). Cardiac myonuclei primarily included the previously distinguished classical myonuclei as well as the recently reported "cytoplasmic myonuclei" (*33*) (fig. S8 and supplementary text note S2). Other myonuclei subsets included neuromuscular junction (NMJ)–localized skeletal muscle myonuclei, which have also been observed in scRNA-seq and snRNA-seq studies in mice (*21*, *31*, *35*), and "fast-twitch" and "slow-twitch" subtypes (Fig. 1, B and C, and fig. S6G), which are characterized by differentially expressed genes (fig. S9, A and B) that are concordant with previously reported markers (*36*) (fig. S9, C and D).

Cross-tissue and cross-sample integration enhanced our ability to resolve multiple rare cell subsets (Fig. 1C and figs. S2, B and D, S5, and S6). For example, we detected Schwann cells that support peripheral nerves (*37*) in multiple tissues (esophagus mucosa and muscularis, heart, prostate, and skeletal muscle), rare neuroendocrine cells in the prostate (*38*), and rare (*26*) ICCs and enteric neurons in the esophagus. Because these rare cells can contribute to various pathologies, their profiling in human tissues will enable disease studies (*26*, *37*).

snRNA-seq protocols perform well across tissues and correspond to scRNA-seq

We benchmarked the performance of our nucleus extraction and profiling protocols (*26*, *29*) relative to each other across all eight profiled tissues and to other snRNA-seq, scRNA-seq, and bulk RNA-seq datasets in relevant tissues. For each dataset, we compared standard quality control metrics per profiled cell or nucleus, as well as the diversity and proportions of captured cell types.

Of the four tested nucleus isolation protocols (CST, NST, TST, and EZ; table S1), the EZ protocol displayed lower performance in each of the eight profiled tissues by multiple quality metrics (*28*, *39*) (Fig. 2A and fig. S10). These included the lowest total number of nuclei captured (fig. S10, A and B), higher levels of ambient RNA (FDR < 0.05; fig. S1, A and B), and separate grouping of EZ-profiled samples [fig. S2C; (*28*)].

The extraction protocols also varied in the proportion of nuclei recovered from each cell type (figs. S2, B and D, and S11A; supplementary text note S3), consistent with our previous observations in tumors (*29*). The TST, CST, and NST protocols had comparable cell-type diversity as measured by Shannon entropy [Fig. 2A; (*28*)], whereas the EZ protocol resulted in significantly lower diversity (Fig. 2A; linear mixed-effects model effect size of −1.08, $P = 5 \times 10^{-11}$). Overall, TST yielded the highest cell-type diversity, on average, across tissues (Fig. 2A) and significantly higher proportions of T cells, fibroblasts, and vascular endothelial cells (FDR < 10%; fig. S11A). Because the protocols varied by their performance (most diverse, high capture of the desired cell types), users should choose protocols by matching protocol features to scientific goal, tissue type, and complexity; and further protocol optimization may still be required (*29*) (for further details and guidance, see supplementary text note S3).

We compared cell-type compositions between our four protocols and other snRNA-seq studies, focusing on our frozen heart left-ventricle samples, for which two recently published snRNA-seq studies evaluated similar samples (*33*, *34*). We found agreement in broad cell types such as mast cells, adipocytes, "cytoplasmic" cardiac myonuclei, and Schwann cells (fig. S11, B and C, and table S4)—but differences in some of their proportions (fig. S11, D and E). Protocols used in the published studies and the EZ protocol in our study captured a higher proportion of muscle nuclei, whereas CST, NST, and TST yielded a higher proportion of endothelial cells (fig. S11E). There was also high concordance between the expression profiles of bulk RNA-seq [from GTEx; (*7*)] and pseudobulk profiles derived from our snRNA-seq (accuracy 92.3%; fig. S12). A few samples showed lower agreement (heart-EZ, breast-EZ, and two breast-TST samples), suggesting that these particular tissue-protocol combinations may not reflect cellular composition as accurately.

We next compared snRNA-seq data to fresh-tissue scRNA-seq data from lung (*40*), skin [current study; (*28*)], and prostate (*38*). For cell composition (Fig. 2B), we recovered the same main cell groups across compartments. We confirmed the accuracy of our annotations by training a multiclass random forest classifier on snRNA-seq data and predicting cell types on scRNA-seq data [(*28*); Fig. 2, C to E], and vice versa (fig. S11, F to H). In addition, cell-type intrinsic (pseudobulk) profiles of protein-coding genes were overall similar between snRNA-seq and scRNA-seq [average Spearman ρ across cell types of 0.58 (skin), 0.69 (prostate), and 0.47 (lung); table S4]. Moreover, integrating cell and nuclei profiles from prostate, skin, and lung and annotating the cells with a random forest classifier that is trained on our nuclei profiles with our granular annotations yielded well-mixed groupings, similar marker genes, and high concordance between protocols (fig. S13).

Divergences observed include the greater expression in cells versus nuclei of a dissociation-induced stress signature (*41*, *42*) (Wilcoxon rank sum test, Benjamini-Hochberg FDR < 10^{-16}; Fig. 2F and fig. S14, A and B), as reported (*29*), and of ribosomal and nuclear-encoded mitochondrial protein genes [Fig. 2G; (*28*); linear model], consistent with their longer half-lives and higher cytoplasmic levels (*43*, *44*). Conversely, nuclei profiles had higher levels of longer transcripts (fig. S14, H and I) and of transcripts with a larger number of adenine stretches (Fig. 2G and fig. S14, C to G), consistent with previous reports (*45*).

Notably, our snRNA-seq generally captured relatively lower proportions of lymphocytes. For example, for lung and skin, respectively, T cells represented 1.7 and 1.4% of all cells (aggregated) compared with 8.73 and 6.83% by scRNA-seq. We observed similar patterns for B cells in skin. Furthermore, these immune cell proportions varied across samples and protocols. A study comparing snRNA-seq and in situ measurements (*46*) suggested that scRNA-seq may oversample immune cells.

Myeloid populations across tissues

Our cross-tissue atlas allowed us to characterize tissue-specific and shared features of

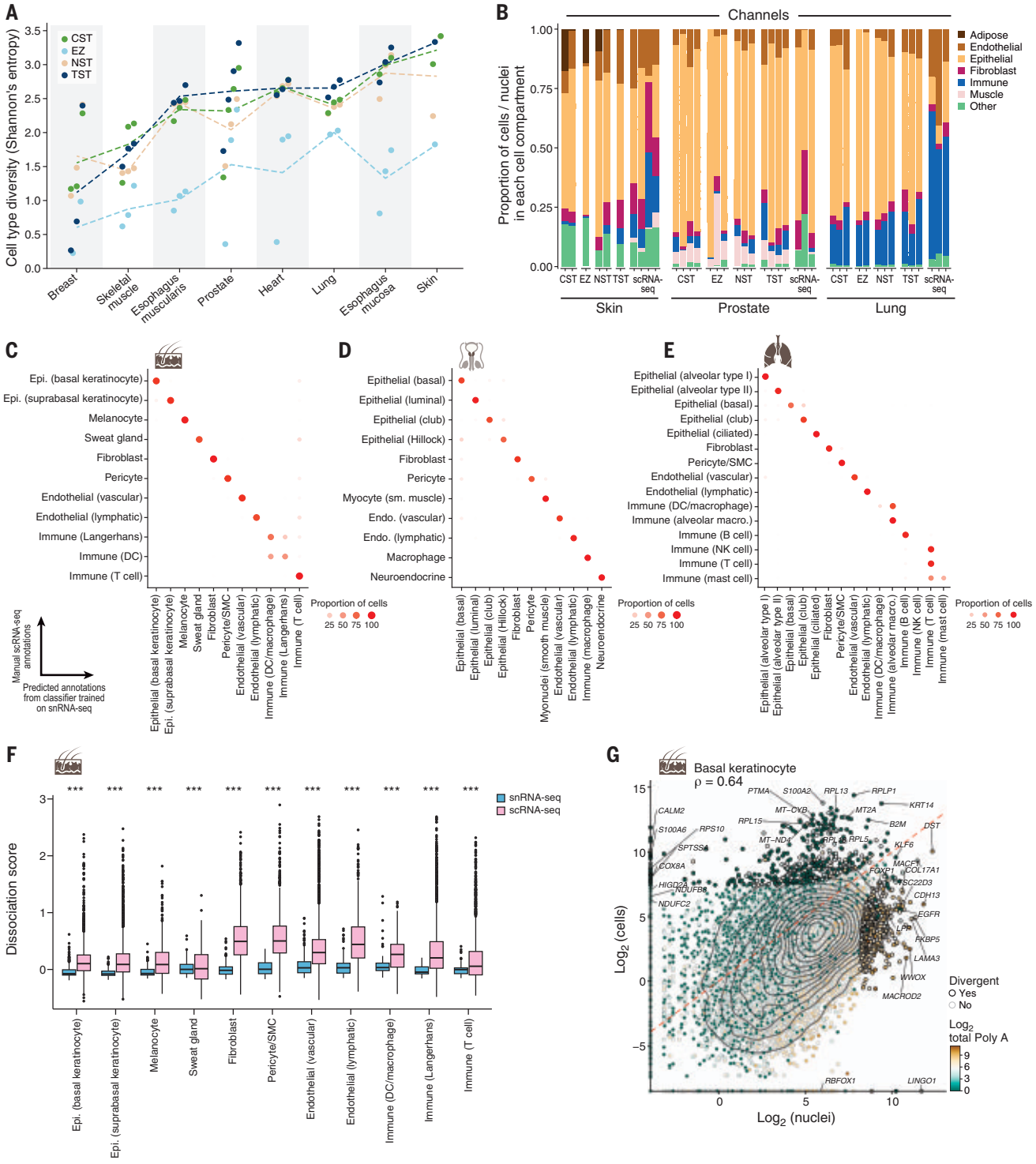


Fig. 2. Concordance of cell-type diversity and cell-intrinsic profiles between snRNA-seq and scRNA-seq. (**A**) Cell-type diversity (Shannon entropy, *y* axis) of each protocol (color) in each sample (dot) and tissue (*x* axis). Dashed lines indicate the average across samples. (**B**) Differences in cell proportions. The proportions (*y* axis) of cells from major categories (color) in each individual by tissue and protocol (*x* axis) are shown. (**C** to **E**) Concordance of cell-intrinsic programs. Proportions of cells (dot color and size) of a manually annotated group (rows) predicted to belong to a given nucleus profile annotation label (columns) by a random forest classifier trained on nuclei and applied to cells of the same tissue for skin (C), lung (D) or prostate (E) are shown. (**F**) Tissue dissociation expression signatures in scRNA-seq. Scores [*y* axis, average background corrected log(TP10K+1)] of a dissociation-related stress signature (*41*) in scRNA-seq (pink) and snRNA-seq (blue) profiles in each major lung cell type (*x* axis) are shown (***Benjamini-Hochberg FDR < 10^{-16}, Wilcoxon rank sum test). Box plots show median, quartiles, and whiskers at 1.5 times the interquartile range (IQR). (**G**) Divergent genes between cell and nucleus profiles. Averaged pseudobulk expression (*28*) of protein-coding genes (dots) in skin basal keratinocyte nuclei (*x* axis) and cells (*y* axis) is shown. Divergent genes are represented by a black dot outline. The color scale shows the total length of polyA stretches with at least 20 adenine bases in $\log_2$ scale. Epi., epithelial; sm., smooth; SMC, smooth muscle cell.

tissue-resident immune cells, which play key roles in immune surveillance and tissue support (*47*, *48*). Integration and annotation of 14,156 myeloid nuclei profiles (*28*) (60% of immune nuclei) revealed 14 distinct monocyte, macrophage, and dendritic cell (DC) subsets (Fig. 3A; fig. S15, A and B; and table S5). These included *CD16*$^{+}$ monocytes, *CD14*$^{+}$ monocytes, two transitional Mo/MΦ *FCGR3A*^{low} and Mo/MΦ *FCGR3A*^{high} populations with coexpression of both monocyte and macrophage markers (see next section), DC1s, DC2s (*49*), mature DCs, and Langerhans cells (Fig. 3B). Tissue macrophage states further included lung macrophages expressing an alveolar macrophage signature (*50*) (fig. S15C), proliferating macrophages, cytokine- and chemokine-expressing inflammatory macrophages, and two additional macrophage subsets: MΦ *LYVE1*^{high} and MΦ *HLA*II^{high} (Fig. 3B), where *HLA*II is *HLA* class II. Finally, lipid-associated macrophage (LAM)–like nuclei highly expressed LAM signature (Fig. 3B and fig. S15C) (*51*) as well as lipid metabolism-related, myeloid cell immune activation, and macrophage migration genes (fig. S15D). Although most myeloid subsets were present in multiple tissues, notable exceptions included *PPARG*^{high} lung macrophages, which were present only in lung, and *CD207*/Langerin^{+} Langerhans cells, which were present only (97%) in skin and esophagus mucosa, consistent with their role in antigen sampling within stratified epithelia (*52*, *53*) (Fig. 3C and fig. S15, F and I).

Myeloid cell proportions were more highly correlated between samples within a tissue type (fig. S15, E and F) than between different tissues (fig. S15, E to H), confirming the reproducibility of tissue-specific myeloid state proportions. Moreover, related tissues—such as muscle (heart, esophagus muscularis, skeletal muscle) or epithelial barriers (esophagus mucosa, skin)—grouped by their myeloid composition profiles (fig. S15H). Macrophage types and proportions varied by tissue, with breast, esophagus mucosa, esophagus muscularis, heart, and skeletal muscle having significantly higher proportions of MΦ *LYVE1*^{high} macrophages [$P < 10^{-6}$, Dirichlet regression likelihood ratio test (LRT); (*28*)] and lung and prostate having significantly higher proportions of MΦ *HLA*II^{high} [$P < 10^{-8}$, Dirichlet regression LRT; (*28*)] (Fig. 3C and fig. S15, F and I).

A dichotomy between LYVE1- and HLAII–expressing macrophages is preserved across tissues

Two expression states of *LYVE1*^{high} and *HLA*II^{high} macrophage populations were dichotomous—either *LYVE1*^{high}*HLA*II^{low} or *LYVE1*^{low}*HLA*II^{high}—and represented the end points of two alternative branches. Specifically, low-dimensional representation of monocytes, macrophages, and Mo/MΦ populations as a continuum with diffusion maps captures the *HLA*II^{high} and *LYVE1*^{high} cells as "terminal" points in two branches that emanate from *CD14*$^{+}$ monocytes at the root. Each of the two terminals is preceded by distinct earlier putative transitional states: a Mo/MΦ *FCGR3A*^{low} state between *CD14*$^{+}$ monocytes and the *LYVE1*^{high}*HLA*II^{low} population and a Mo/MΦ *FCGR3A*^{high} state between the monocytes and the *LYVE1*^{low} *HLA*II^{high} cells (Fig. 3D and fig. S16, A and B). (There is also a putative secondary path between *FCGR3A*^{low} and *LYVE1*^{low}*HLA*II^{high} cells through a *FCGR3A*^{high} intermediate.) The position of the *FCGR3A*^{high} transitional state is consistent with lung data from a humanized mouse model (*54*, *55*). These key features are consistent overall in the map that is constructed only from macrophage nuclei from a single tissue (fig. S16A). Thus, *FCGR3A*^{low} and *FCGR3A*^{high} states might be Mo/MΦ populations that are less-differentiated or less-activated states of *LYVE1*^{high} and *HLA*II^{high} MΦs, respectively.

Each of the *LYVE1*^{high} and *HLA*II^{high} subsets expressed a combination of a common signature and tissue-specific markers (Fig. 3E, fig. S16C, and table S5) and was enriched for distinct functions, mirroring those of *Lyve1*^{high} MHCII^{low} and *Lyve1*^{low}MHCII^{high} resident macrophage populations in mouse tissues (*47*) (fig. S16, C and E). *HLA*II^{high} cells were enriched for immune-related processes and expressed higher levels of complement components *APOE* and *C1QA*, *C1QB*, and *C1QC* (Fig. 3B and fig. S16, C and D). Genes differentially expressed between human *HLA*II^{high} and *LYVE1*^{high} subsets corresponded to those in murine counterparts (fig. S16E), but with higher expression in human *HLA*II^{high} macrophages of *C1QB* and *C1QC* complement genes in lung and phagocytic receptors *MARCO* and *CD36* in heart (fig. S16E). *HLA*II^{high} macrophages were also enriched for immune interactions with B cells, DCs, mast cells, natural killer (NK) cells, and T cells (fig. S16G). *LYVE1*^{high} profiles were enriched for tissue-supporting modules and had putative receptor-ligand interactions (*28*) with lymphatic endothelial cells, fibroblasts, adipocytes, and myocytes (fig. S16F). In mice, MΦ *Lyve1*^{high} cells were located near blood vessels (*47*) and regulated vascular tone (*56*), and cardiac *LYVE1*$^{+}$ macrophages have been implicated in regulating the lymphatic network (*57*). Thus, *LYVE1*^{high} macrophages may have a homeostatic role in the human heart, lung, and esophagus.

LAM-like macrophages are prevalent across human tissues and share a regulatory program

In our atlas, we identified LAM-like cells as widely distributed across healthy human tissues, with the vast majority of LAMs (97%, 268 of 283) from the breast, heart, lung, and prostate (Fig. 3C). LAMs and LAM-like cells have been previously reported in disease contexts in adipose tissue from obese humans and mice (*51*), injured and fibrotic liver (*58*–*60*), obese liver (*61*), fibrotic lung (*50*, *62*), atherosclerotic aortic tissue (*63*, *64*), leprosy (*65*), and the brains of individuals with Alzheimer's disease (*66*–*68*). However, an understanding of their distribution and heterogeneity across human tissues is still lacking.

To characterize LAMs across the body, we analyzed LAM-like cells in the expanded context of our study and 17 other published atlases spanning 14 tissues. We trained a linear classifier with published omental adipose scRNA-seq containing LAMs (*28*, *51*) and classified each myeloid profile in our dataset and the published compendium as LAM-like macrophages, non-LAM macrophages, and non-macrophages. From this, we recovered 283 LAM-like cells in our study and 4285 LAM-like cells in the 17 published studies (Fig. 3, F to H; fig. S17, A to D; and table S6).

LAM-like cells were present among tissue-resident macrophages across a broad range of tissues and pathologies. These included adipose (*51*, *69*) and atherosclerotic (*70*, *71*) tissue, as reported, as well as healthy tissues [placenta (*72*), testis (*73*), kidney (*74*), pancreas (*75*), prostate (*38*), decidua (*72*), liver (*76*), ovary (*76*, *77*), skeletal muscle (*78*), and intestine (*26*)] and other disease contexts [acne (*79*), leprosy (*79*) and atopic dermatitis (*80*, *81*) (skin), and Crohn's disease (ileum) (*82*)]. Microglia from the central nervous system of epileptic patients (*83*) were also classified as LAM-like cells, indicating that microglia that express LAM signature genes extend beyond Alzheimer's disease (*66*–*68*). LAM signature genes were enriched for genome-wide association study (GWAS) genes associated with levels of high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterol, triglycerides, type 2 diabetes, and the tau to Aβ1-42 ratio in cerebrospinal fluid (fig. S17H and table S7), further supporting their role in lipid homeostasis.

Although a core set of signature LAM genes was expressed across most tissues and studies, many other genes varied across tissues. For example, both *CHIT1* and *CTSK* were highly expressed in LAMs from leprosy skin samples (but had very low expression in other skin LAMs) (Fig. 3H and fig. S17F), in line with higher serum chitotriosidase activity in leprosy patients (*84*). Select lipid pathway genes, including fatty-acid binding protein *FABP4*, lipoprotein lipase *LPL*, and phagocytic lipid receptor *CD36*, were highly expressed in LAMs from tissues with high adipose content, including adipose, atherosclerotic lesions, and intestine creeping fat samples, possibly reflecting increased lipid-induced transcriptional stimulation of these target genes under these conditions (*85*). LAMs in creeping fat from

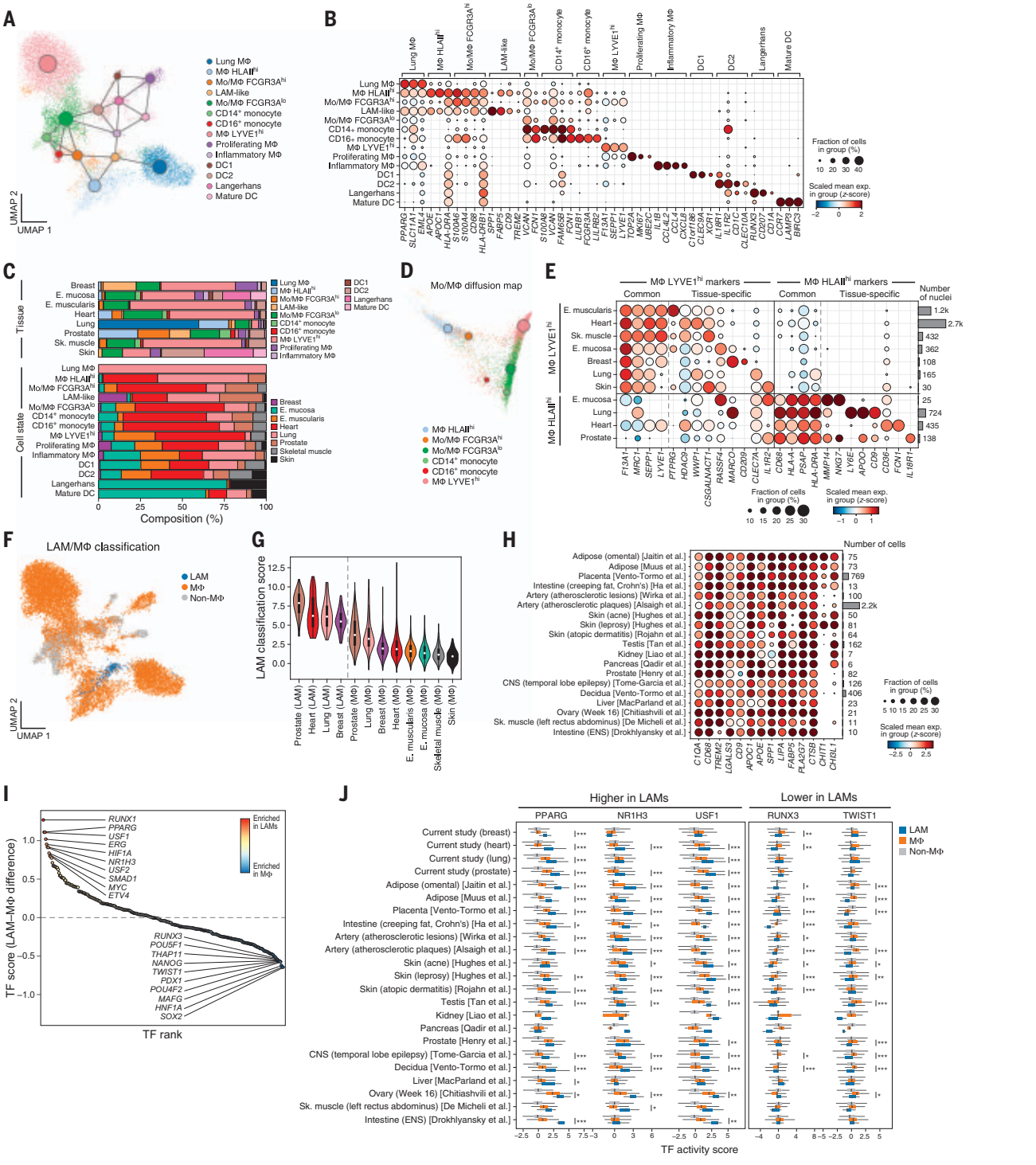


Fig. 3. A dichotomy between *LYVE1*- and *HLA*II–expressing macrophages and LAM-like populations across tissues. (**A**) Myeloid profiles (dots), colored by cell type and state and overlaid with a PAGA graph of myeloid states (large nodes). (**B**) Expression of marker genes (columns) associated with each subset (rows). (**C**) Myeloid cell distribution across tissues. The overall proportion of myeloid cell subsets (colors) in each tissue (bars) is shown at the top, and the overall proportion of cells from each tissue in each subset (bars) is shown at the bottom. (**D**) *LYVE1*^{high} and *HLA*II^{high} macrophages are end points of two differentiation trajectories. A diffusion map of monocytes, macrophages, and transitional subsets (colors) is shown. Large circles represent centroids (sizes are proportional to population size). (**E**) Cross-tissue and tissue-specific markers. Expression of marker genes (columns) associated with two myeloid subsets (left) in each tissue (rows) is shown. The right bar plot shows the number of nuclei. (**F** to **H**) LAM-like cells across tissues. Myeloid cells (dots) colored by their classification [legend; (*28*)] are shown in (F). Classification scores (*y* axis) of LAM-like and other macrophages across tissues (*x* axis) are shown in (G). Expression of LAM marker genes (columns) in LAM-like profiles from other studies (rows) is shown in (H). (**I** and **J**) Inferred TFs regulating the LAM-like program. TF differential activity scores between LAMs and other macrophages (*y* axis) for each TF (dot) ranked by score (*x* axis) are shown in (I). TF differential activity scores (*x* axis) for three TFs with significantly high scores (two tailed *t* test; Benjamini-Hochberg *FDR < 0.05, **FDR < 0.01, and ***FDR < 0.001) in LAMs or other macrophages are shown in (J). Box plots show median, quartiles, and whiskers at 1.5 times the IQR. E., esophagus; ENS, enteric nervous system; Sk., skeletal.

Crohn's intestine and atherosclerotic lesions were additionally characterized by higher expression of interleukin *IL1B* and multiple chemokines (*CXCL8*, *CXCL3*, *CCL4*), possibly reflecting the inflammatory environment under these conditions (fig. S17F).

We predicted transcription factors (TFs) that could mediate LAM-like gene expression by inferring TF activity from target expression (*28*) and ranking TFs by the mean difference between their activities in LAMs versus non-LAM macrophages [(*28*); Fig. 3I]. LAM-associated TFs inferred across all classified LAM-like cells included *PPARG*, *USF1*, and *NR1H3* (*LXRA*) (Fig. 3J and fig. S17G), suggesting a shared core regulatory mechanism. These are major regulators of lipid metabolism–related expression (*86*) and have been proposed to regulate *Trem2* expression in mice (*87*).

Shared and tissue-specific features of fibroblasts

To characterize fibroblast heterogeneity (*88*), we analyzed 32,421 fibroblast nuclei profiles across the eight profiled tissues (fig. S18A), identifying shared and tissue-specific signatures. The cross-tissue, shared fibroblast expression program consisted of markers that were significantly more highly expressed in fibroblasts than in nonfibroblast cell types within each tissue (FDR < 0.05, Welch's *t* test) and included multiple extracellular matrix (ECM) constituents (Fig. 4A and table S8). Conversely, the tissue-enriched fibroblast signatures were defined based on genes exclusive to or highly enriched in fibroblasts from a given tissue versus fibroblasts from all other tissues (Fig. 4, B to D, and fig. S18B).

Tissue-enriched fibroblast features were consistent with the specific functions and interactions required in the respective tissues. For example, the esophagus mucosa fibroblast signature (table S8) included genes involved in neuron and axon development (e.g., *NTN1*, *PLXNB1*, *FGF13*), suggesting interactions with the enteric nervous system, possibly through NTN1-DCC and NTN1-UNC5C interactions (fig. S18E). The cardiac fibroblast signature genes included TFs involved in cardiac development (e.g., *GATA4* and *GATA6*), revealing that expression of these developmental TFs is retained in the adult cardiac fibroblast compartment (*89*–*92*). The skeletal muscle signature showed increased expression of the *CXCL14* and *CXCL12* chemokines and of components of the renin-angiotensin-aldosterone system (*AGTR1* and *MME*), which regulate skeletal muscle mass (*93*, *94*) [reminiscent of the expression of the same components by WNT2B$^+$ fibroblasts in a local renin-angiotensin system in the colon (*16*)].

Lung fibroblast signatures were enriched for ECM, cation transport, and contractile functions (Fig. 4E), including multiple components of the basement membrane (BM) that are required for epithelial-mesenchymal adhesion [nephronectin (*NPNT*), *FRAS1*, hemicentin-1 (*HMCN1*), and integrin *ITGA8* (Fig. 4, F and G), which form a protein complex (*95*–*98*) that anchors NPNT to the BM (*99*–*103*)]. Mutations in *ITGA8*, *FRAS1*, and *HMCN1* have been linked to Fraser syndrome, a congenital disorder that affects cell adhesion and results in skin, kidney, lung, and craniofacial abnormalities (*104*–*107*), and common variants in *NPNT* are significantly associated with chronic obstructive pulmonary disease (COPD) and forced expiratory volume (FEV) in GWASs (Fig. 4H). Granular annotation of the lung fibroblast profiles (*40*, *108*) (fig. S18C) suggests that the adhesion complex is expressed by alveolar fibroblasts (Fig. 4F and fig. S18D). Alveolar fibroblasts also specifically expressed *FGFR4* (Fig. 4B), which is essential for alveologenesis in mice (*109*) and is genetically linked with bronchopulmonary dysplasia that affects alveoli in humans (*110*).

Of all tissue fibroblasts, lung alveolar fibroblasts distinctively expressed a calcium signaling and actomyosin contractility program (Fig. 4, E and F, and fig. S18D), including the mechanosensitive calcium ion channel *PIEZO2* (Fig. 4G), which has been proposed to sense pulmonary stretch (*111*, *112*); multiple calcium channels and adrenergic and purinergic receptors, which have been implicated in stimulating calcium release from intracellular stores (*CACNA1D*, *TRPC6*, *MCOLN2*/*TRPML2*); and myosin light chain kinase (*MYLK*), which is involved in mediating calcium-induced actomyosin contraction (*113*). This suggests that alveolar fibroblasts could integrate mechanical stretch and forces (through *PIEZO2*) and neuronal excitatory signals (*40*), which could affect their migratory or mechanical properties. Notably, although alveolar fibroblasts express myofibroblast markers (e.g., *ACTA2*; fig. S18D), they are distinct from previously reported myofibroblasts (*40*), and our data suggest that they constitute a distinctive contractile and excitable fibroblast state.

Intra- and cross-tissue cell-type associations with monogenic disorders

Human genetics has identified numerous rare monogenic disease genes, and many have been experimentally mapped to cell type(s) of action (*114*, *115*). To characterize the expression of monogenic disease genes across cell types, we related disease genes associated with different phenotypes or disease categories in the Online Mendelian Inheritance in Man (OMIM) database (*116*) to the cell populations in which they are expressed in our cross-tissue atlas. Because OMIM entries are not organized by disease categories, we leveraged topic modeling to aggregate 5812 genotype-phenotype associations based on similarities in text descriptions of clinical features, resulting in 229 distinct disease topics [(*28*); figs. S19 and S20). We then related topics to a cell type based on the enriched expression of the topic genes in the cell type (Fig. 5A, figs. S21 and S22, and table S9).

Many topics mapped to their expected cell populations. For example, cardiac disease topics (topics 65 and 66) mapped to cardiac myonuclei and/or endothelial cells, immune and infection topics (topics 132, 205, and 217) mapped to immune cells across tissues, and a diabetes and lipodystrophy topic (topic 222) mapped to adipocytes in skeletal muscle and skin (Fig. 5A and fig. S23A). Male infertility and sperm motility topic (topic 129) was associated with ciliated lung cells and stromal cells (fig. S23B). The link between sperm dysmotility and bronchitis is well established, with genes common to flagella and cilia perturbed across both sperm and lung (*117*); we further detect this specific link in topic 155, which associates ciliary dyskinesias and the lung ciliated epithelium (fig. S23C).

Genes from monogenic muscle disease groups are enriched in distinct subsets of myocyte and nonmyocyte nuclei in three muscle types

Among the monogenic disorders, muscle disease phenotypes are a well characterized subset and are known to arise from mutations in genes that are expressed in myocytes (e.g., structural genes involved in contraction) and/or other cells in the surrounding tissue (e.g., NMJ, ECM, and adipose tissue) (*118*–*121*). We leveraged the three muscle types represented in our atlas—cardiac, skeletal, and smooth muscle—to map 605 well-curated monogenic muscle disease genes (*118*) (table S10), recovering known biology and extending hypotheses beyond those obtained from bulk-tissue RNA-seq (*122*, *123*). We tested disease groups (e.g., hereditary cardiomyopathies, motor neuron diseases) for their enrichment with cell type–specific markers across our muscle tissues (FDR < 0.1) [(*28*); Fig. 5B, fig. S24, and table S11].

As expected, different disease gene sets were associated with different myonuclei subsets (113 of 605 genes; table S11) in patterns that recapitulated known disease mechanisms (Fig. 5B). For example, skeletal muscle myonuclei were associated with congenital myopathy genes (FDR = 7.07×10^{-5}), and cardiac myonuclei were associated with hereditary cardiomyopathy genes (FDR = 4.24×10^{-12}). Some associations highlighted finer myonuclei subsets. For example, genes linked to congenital myasthenic syndrome, a disorder affecting neuromuscular transmission (*124*), were specifically expressed in NMJ-localized myonuclei but not in other skeletal myonuclei (tables S2 and S12). These included acetylcholine receptor subunits (*CHRNE*, *CHRNA1*, *CHRND*), the NMJ-organizing receptor tyrosine kinase *MUSK*,

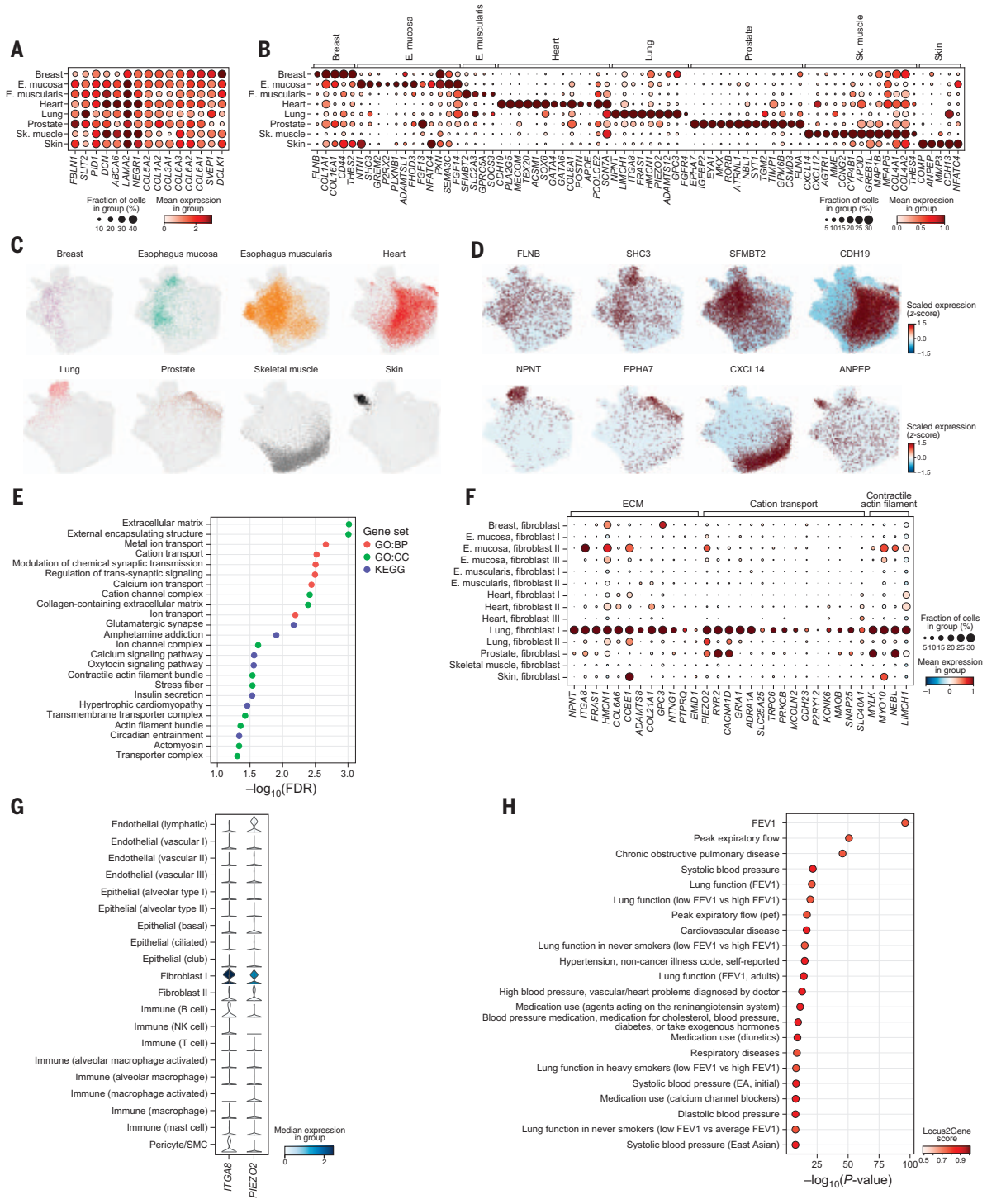
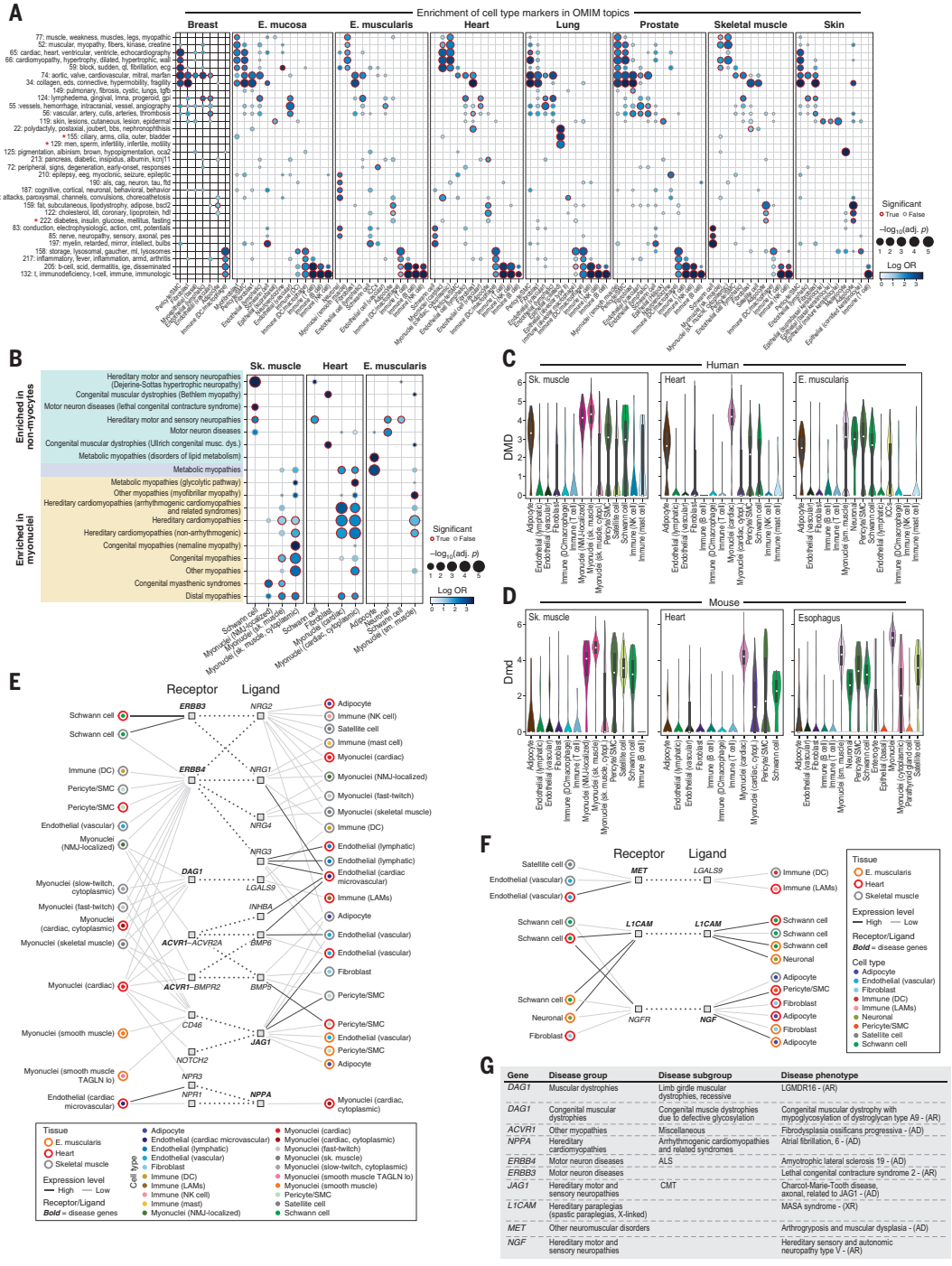


Fig. 4. Shared and tissue-specific fibroblast features. (**A** and **B**) Expression in each tissue subset (rows) of marker genes (columns) distinguishing fibroblasts from nonfibroblasts across all tissues (A) or enriched in fibroblasts in one versus other tissues (B). (**C** and **D**) Fibroblast profiles (dots) colored by tissue (C) or expression of the most exclusive marker (D). (**E**) Significance [$-\log_{10}$(FDR), *x* axis] of gene sets (*y* axis) enriched (FDR < 5%) in genes covarying with the lung-specific fibroblast signature. (**F**) Expression of ECM and cation transport genes (columns) in the covarying gene module in each granular fibroblast subtype in each tissue (rows). (**G**) *ITGA8* and *PIEZO2* (columns) expression in granular cell types (rows) in lung. (**H**) Significance (*x* axis) and Open Targets Genetics locus-to-gene score (color) of the most significant variants mapped to *NPNT* with a high (>0.5) locus-to-gene score in GWASs (*y* axis). FEV, forced expiratory volume.

and NMJ-enriched ECM components (*COL13A1*, *LAMA2*).

Other disease gene sets were associated with nonmyocyte accessory cells, including neurons, Schwann cells, fibroblasts, and adipocytes (127 genes; table S12), often mirroring clinical features, such as nervous system cells in neuropathies or adipocytes in metabolic myopathies. In particular, Schwann cells were associated with hereditary motor and sensory neuropathies in all three tissues (FDR = 0.015 to 0.06), but their association with Dejerine-Sottas



Gene	Disease group	Disease subgroup	Disease phenotype
DAG1	Muscular dystrophies	Limb girdle muscular dystrophies, recessive	LGMDR16 - (AR)
DAG1	Congenital muscular dystrophies	Congenital muscle dystrophies due to defective glycosylation	Congenital muscular dystrophy with myoglycosylation of dystroglycan type A9 - (AR)
ACVR1	Other myopathies	Miscellaneous	Fibrodysplasia ossificans progressiva - (AD)
NPPA	Hereditary cardiomyopathies	Arrhythmogenic cardiomyopathies and related syndromes	Atrial fibrillation, 6 - (AD)
ERBB4	Motor neuron diseases	ALS	Amyotrophic lateral sclerosis 19 - (AD)
ERBB3	Motor neuron diseases		Lethal congenital contracture syndrome 2 - (AR)
JAG1	Hereditary motor and sensory neuropathies	CMT	Charcot-Marie-Tooth disease, axonal, related to JAG1 - (AD)
L1CAM	Hereditary paraplegias (spastic paraplegias, X-linked)		MASA syndrome - (XR)
MET	Other neuromuscular disorders		Arthrogryposis and muscular dysplasia - (AD)
NGF	Hereditary motor and sensory neuropathies		Hereditary sensory and autonomic neuropathy type V - (AR)

Fig. 5. Monogenic muscle disease genes related to cell types and interactions across cardiac, skeletal, and smooth muscle tissues. (**A**) Enrichment of monogenic disease groups to broad cell types. Effect size (log odds ratio, dot color) and significance [−log$_{10}$(FDR), dot size] of enrichment of genes from disease topics [rows; (*28*)] in broad cell-type markers in each tissue (columns) are shown. A red outline indicates an FDR less than 0.1. Topic names consist of the topic identifier and five words with the highest loadings. Red stars indicate highlighted topics. (**B**) Relation of broad cell types to monogenic muscle disease groups. Effect size and significance of enrichment of genes from monogenic muscle disease groups (rows) for broad cell type markers in each tissue (columns) are shown. A red outline indicates an FDR less than 0.1. Color shading indicates disease groups associated with only nonmyocytes (green), only myonuclei (yellow), or both (light purple). (**C** and **D**) *DMD* expression in human (C) and mouse (D) muscle. Cell types (*x* axis) are ordered, left to right, such that the cell types that are shared between human and mouse within a tissue are presented first and species-specific cell types follow. (**E** and **F**) Putative cell-cell interactions in muscle implicating muscle disease genes. Shown are cell types (inner color) from muscle tissues (outer color) connected by putative interactions (dotted edges) between a receptor (left square) expressed in one cell type and a ligand (right square) expressed in the other in interactions involving myocytes (E) or only nonmyocytes (F). Black and gray connecting lines between cell types and genes indicate high and low expression, respectively. Bold formatting indicates a muscle disease gene. (**G**) Diseases highlighted in (E) and (F). ALS, amyotrophic lateral sclerosis; AD, autosomal dominant; AR, autosomal recessive; CMT, Charcot-Marie-Tooth disease; XR, X-linked recessive.

hypertrophic neuropathy (a subtype of Charcot-Marie-Tooth disease) was specific to skeletal muscle (FDR = 2.28×10^{-5}), consistent with the need to maintain innervation to prevent muscle atrophy (*125*). Thus, distinctive cell expression patterns can provide insight into onset (Dejerine-Sottas is an early-onset disease), severity, and the predominant peripheral nerves that are affected (i.e., motor, sensory, or autonomic), where heart can possibly act as a proxy for autonomic nerves. In other examples, adipocytes from esophagus muscularis were associated with metabolic myopathies related to lipid metabolism (FDR = 0.0003); the congenital muscular dystrophies genes *COL6A1*, *COL6A2*, and *COL6A3* are expressed in fibroblasts, consistent with joint laxity and progressive contractures along with muscle weakness (*126*); and the recessive limb girdle muscular dystrophies gene *DYSF* is expressed in immune cells, consistent with the role of more aggressive monocytes in disease progression (*127*). Some of the enrichments in nonmyocytes are also present in the same cell types in other nonmuscle tissues (e.g., breast adipocytes for metabolic myopathies or breast and skin pericytes for cardiomyopathies; fig. S25I), highlighting that tissue-specific pathology may arise from the relation between an accessory cell's broader function and specialized tissue demand.

There were differences in the expression of monogenic muscle disease genes between slow and fast myocyte subsets for both classical and cytoplasmic myonuclei (fig. S9). As expected, slow myocytes preferentially expressed type 1 fiber markers (*36*, *128*) and disease genes *MYH7* (FDR < 10^{-15} in cytoplasmic myonuclei) and *PPARGC1A* (FDR < 10^{-15} in regular myonuclei). Fast myocytes specifically expressed the type 2 fiber marker *MYH2* (FDR < 10^{-15} in cytoplasmic myonuclei), in line with respective clinical myosinopathy phenotypes (*129*). Congenital myopathy disease genes, which include genes known to be active in slow and fast myocytes, were enriched in both groups. The myotonic dystrophy (MD) type 2 gene *CNBP*, was enriched in regular myonuclei from fast myocytes [FDR < 10^{-12}; (*28*); fig. S9, E and F], which are preferentially affected in MD type 2 (*130*, *131*). By contrast, MD type 1 disease gene *DMPK* was preferentially expressed in cytoplasmic myonuclei from slow myocytes ($P < 0.036$, Welch's t test), which are perturbed in MD type 1 (*132*). Regular myonuclei subsets from slow myocytes were also enriched in *TPM3* [FDR < 10^{-15}; (*28*); fig. S9, E and F], which has been linked to congenital fiber type disproportion, a condition characterized by smaller slow myocyte fibers (*133*). Metabolic myopathy disease genes (*GBE1*, *ETFDH*, and *SLC25A20*) were also enriched in slow myocyte markers ($P < 0.05$, Fisher's exact test; fig. S9, E and F).

Many of the cell-type associations for monogenic muscle disease genes were conserved between our human muscle atlas and corresponding mouse snRNA-seq data (*26*, *28*) (fig. S25), but there were also notable differences. In both mouse and human, there were significant associations between skeletal muscle myonuclei and various dystrophies and myopathies (FDR < 0.1, Fisher's exact test), between cardiac myonuclei and cardiomyopathies (FDR < 0.1), between adipocytes (in skeletal muscle, esophagus, and heart) and metabolic myopathies (FDR < 0.1), and between Schwann cells in skeletal muscle and hereditary motor and sensory neuropathies (FDR < 0.1) (fig. S25H). However, dystrophin (*DMD*) expression across accessory cell types varied between human and mouse. In humans, high levels of *DMD* expression (comparable to that in myonuclei) were observed in adipocytes in all muscle types; pericytes and Schwann cells in skeletal muscle and esophagus muscularis; and enteric neurons in esophagus muscularis [mean expression log(TP10K+1) > 2.0; Fig. 5C]. Lower, intermediate levels of *DMD* expression were observed in skeletal muscle fibroblasts and satellite cells, as well as esophagus muscularis ICCs [mean expression log(TP10K+1) < 2.0; Fig. 5C]. *DMD* expression in adipocytes supports the possibility of local metabolic perturbation (*134*) (Fig. 5C), whereas its expression in the enteric nervous system and ICCs raises the possibility that perturbation of these cells contributes to the gastrointestinal dysfunction phenotype in Duchenne muscular dystrophy (*135*). In mouse, whereas *Dmd* expression is high in myonuclei, pericytes, satellite cells, and Schwann cells, it is low in adipocytes and fibroblasts across all three muscle tissue types [mean expression log(TP10K+1) < 1.0; Fig. 5D]. These differences suggest that the effects of *DMD* mutation on accessory cells and their contribution to disease (*135*, *136*) may differ between human patients and mouse disease models (*137*).

Cell-composition and cell-intrinsic secondary effects in muscle disease tissue

Muscle dystrophies commonly display a secondary shift in cellular composition, with fibrotic or adipogenic replacement of muscle tissue (*138*), which can introduce secondary pathologic processes that exacerbate muscle loss (*139*). To characterize the cellular expression patterns of disease genes whose bulk-tissue expression levels are altered in muscle diseases, we analyzed the cell-type specificity of genes up-regulated in bulk RNA-seq data from muscle tissues from 43 patients with rare muscle disorders (*122*) compared with healthy muscle tissues (*140*) (fig. S26). Some of the cell type–specific expression patterns reflected known biology, including up-regulation of hallmark fibroblast, adipocyte, and immune genes, likely because of changes in cell proportions (e.g., *COL6A2*, *ADIPOQ*, and *HLA-A*, respectively). Other patterns suggest additional, cell-intrinsic regulatory events (beyond cell composition changes) that may modify disease progression. For example, dermatopontin (*DPT*), an adipokine that promotes ECM remodeling and inflammation (*141*), is expressed in adipocytes but is not correlated with *ADIPOQ* expression across adipocytes from different tissues. This suggests that its expression is regulated and is not merely reflecting changes in cell composition. Similarly, *ELK3*, which is expressed in endothelial cells, suppresses angiogenesis (*142*) and may contribute to functional muscle ischemia in muscle disease (*143*), which is otherwise typically attributed to nitric oxide signaling (*136*). Thus, the increased resolution of a single-cell atlas can help disentangle secondary effects related to cell-composition and cell-regulatory events in accessory cells during disease.

Disease genes may affect receptor-ligand interactions

Some disease genes encode receptors or ligands that participate in cell-cell interactions, such that loss-of-function mutations can affect tissue function through non–cell autonomous effects through these cell-cell interactions. We related cell types in muscle tissue to one another through receptor-ligand interactions (*16*, *26*, *144*) that included at least one monogenic disease gene (*28*) in every tissue in our atlas (table S13).

Our analysis suggests that mutations in some disease-causing genes may disrupt interactions between myocytes and other cell types. For example, mutations in *ERBB3* (the disease gene for lethal congenital contracture syndrome) may disrupt interactions between myocytes and Schwann cells (Fig. 5E and table S13) and contribute to joint contractures as an associated, but not primary, disease phenotype (*145*). Mutations in *DAG1* (the disease gene for congenital muscular dystrophy), although known to interact primarily with laminin produced by fibroblasts (*146*, *147*), may additionally disrupt interactions with immune cells through *LGALS9* and alter their function (*148*). Putative cell-cell interactions involving only nonmyocytes included the disease genes *L1CAM* (MASA syndrome), *MET* (arthrogryposis and muscular dysplasia), and *NGF* (hereditary sensory and autonomic neuropathy), each potentially affecting multiple cell pairs, including neurons and Schwann, satellite, immune, and stromal cells (Fig. 5, F and G, and table S13).

Cell type–specific enrichment of QTL genes mapped to GWAS loci

Single-cell atlases can also provide insights into cell-type specificity and mechanisms of action of the genes in disease-associated loci identified by GWASs. Studies associating genetic variants to changes in gene expression or splicing

quantitative trait loci (eQTL or sQTL, respectively) showed tissue-specific colocalization with multiple loci from GWASs of human traits, including disease risk (*7*, *149*–*151*), but lacked cellular resolution (*11*). To prioritize cell types of action and causal genes for complex diseases and traits in specific cells and tissues, we used ECLIPSER (*28*, *152*) to test whether GWAS loci from 21 complex traits (table S14), with likely effects in at least one of the eight tissues analyzed, are enriched for genes with high cell type–specific expression in each tissue. We defined putative causal genes for each GWAS locus as the set of genes whose eQTLs and sQTLs (*7*) were in linkage disequilibrium [squared genotype correlation (r^2) > 0.8] with the lead GWAS variant(s) [Fig. 6A; (*28*)]. We further included genes prioritized by additional genomic data [e.g., Hi-C and protein QTLs (pQTLs)] and linkage to predicted deleterious protein-coding variants (*153*, *154*). Because more than one gene typically maps to a GWAS locus using this approach (mean = 2, and maximum = 37 for selected traits and 170 for null traits), we scored loci by the fraction of cell type–specific genes in the locus [Fig. 6A; (*28*)]. We assessed enrichment for each GWAS locus set against a null distribution of GWAS loci associated with tissue-unrelated traits using a Fisher's exact test [Fig. 6A; (*28*)].

Seventeen of the traits were enriched in both expected and previously undescribed cell types at a tissue-wide FDR less than 0.05 (Benjamini-Hochberg), 16 of which were significant (FDR < 0.05) across tissues (Fig. 6B, figs. S27 and S28, and tables S15 and S16). Among the expected associations are skin pigmentation traits in melanocytes, autoimmune and inflammatory diseases in T and NK cells, COPD in lung fibroblasts, prostate cancer in luminal epithelial cells, atrial fibrillation and heart rate in myonuclei, and heart rate in lymphatic endothelial cells (*155*) (Fig. 6B). Type 2 diabetes loci were enriched in skeletal muscle adipocytes and in lymphatic endothelial cells in multiple tissues, which might contribute to the predisposition of type 2 diabetes to vascular disease (*156*, *157*) (Fig. 6B). Less well-characterized cell type–trait associations included DCs (in almost all tissues) with non-melanoma skin cancer (*158*) and adipocytes (breast) with atrial fibrillation (table S15). GWAS loci enriched in a specific cell type from a known tissue of action frequently showed similar enrichment in the same cell type from other uninvolved tissues. For example, atrial fibrillation GWAS loci were enriched in myonuclei in heart, skeletal muscle, esophagus muscularis, and prostate (Fig. 6, B and C); coronary artery disease and heart rate loci were enriched in pericytes in five or six tissues in addition to heart; and prostate cancer loci were enriched in luminal epithelial cells in both prostate and breast (Fig. 6B and figs. S29 and S30).

Cell-type enrichment helped identify putative causal genes in GWAS loci with multiple QTL-mapped genes (tables S15 and S16). On average, about two-thirds of genes driving the cell type–specific enrichment for a given trait in a relevant tissue [mean = 66%, 60 to 71.4% (95% confidence interval)] were also driving the enrichment in the same cell type in other tissues. For example, in the case of atrial fibrillation and cardiac myonuclei, 26 out of 31 myonuclei-specific genes that drove the enrichment signal in myonuclei in heart and were shared with at least one other tissue (Fig. 6D) were enriched in muscle system–related processes, such as muscle contraction (FDR < 0.05; table S17). The five myonuclei-specific genes that were specific to heart only (pink vertical lines in Fig. 6D) were instead enriched in heart development processes, such as cardiac muscle tissue development (FDR < 0.05; table S18). Cardiac myonuclei were also found to be the most relevant cell type for atrial fibrillation in two separate snRNA-seq studies of the human heart (*33*, *34*) and based on ECLIPSER analysis of these two studies [fig. S31 and table S19; (*28*)]. *CASQ2*, which encodes a cardiac muscle member of the calsequestrin family, and the myosin heavy chain 6 and 7 genes (*MYH6* and *MYH7*) were the top myonuclei-specific genes driving the enrichment signal for atrial fibrillation in all three studies (table S19).

Associating GWAS genes with gene programs across cell types reveals six main trait groups

To chart cellular programs and processes that may be affected by genetic variants, we associated a larger set of >2000 complex phenotypes with cell types and the covarying gene modules that these cell types express (*28*). We defined gene modules in the cells in our atlas by hierarchically clustering genes based on correlation across all cells, as well as within cell types. We then scored modules for their overlap with GWAS genes [defined by variant to gene mapping in Open Targets Genetics (*153*, *154*)] that are also highly expressed by cell type [(*28*); fig. S32 and table S20]. Next, we grouped GWAS phenotypes into major groups by the similarity of their module enrichment across cell types (Fig. 7, A and B, and fig. S33). Finally, for each major group of traits, we identified the relevant cell types associated with the underlying modules (Fig. 7, A and C) and tested the GWAS genes that overlapped with gene modules for functional enrichments (Fig. 7, A, D, and E).

Traits and diseases partitioned into six major groups, spanning immune hypersensitivity, cardiovascular, calcium channel–related, cognitive and psychiatric, pigmentation, and HDL cholesterol–related based on their associations with cell types (Fig. 7B). The immune hypersensitivity disorders were associated with T cells, including the expected relation between lung T cells and hay fever, allergic rhinitis, and respiratory disease [(*159*, *160*); Fig. 7C]. GWAS genes in the modules associated with these traits were enriched for lymphocyte activation and differentiation and T cell receptor signaling and cell-cell adhesion (Fig. 7D), with interleukin-35 (IL-35) signaling genes enriched in hypothyroidism and inflammatory bowel disease (IBD), consistent with IL-35 up-regulation in Hashimoto's thyroiditis (*161*) and IBD (*162*) and down-regulation in Graves' disease (*163*). The cardiovascular traits group was associated with pericytes and smooth muscle cells and enriched with blood circulation, smooth muscle contraction, muscle structure development, and cardiocyte differentiation genes. The cardiovascular group overlapped with the calcium channel–related group—which included blood pressure medication, pulse rate, medication use of calcium-channel blockers, and vascular system traits, as well as schizophrenia and autism spectrum disorder, which are psychiatric disorders with known calcium channel associations (*164*)—and was enriched with membrane depolarization and calcium ion channel genes (Fig. 7D). Other cognitive and psychiatric phenotypes grouped separately and were enriched with neuronal synapse organization, structure, and activity genes (Fig. 7D) Finally, a group of HDL cholesterol traits was associated with adipocytes in all three muscle tissues and enriched for fatty acid, triglyceride, and lipid homeostasis and related metabolic processes (monocarboxylic and glycerol metabolism), including *ANGPTL8* and *PNPLA3* (Fig. 7E and table S20), which are genes involved in lipolysis regulation in adipocytes that might affect extrahepatic cholesterol transport via HDLs (*165*, *166*).

Toward large-scale snRNA-seq of human tissues with pooling

To enable future studies at the population scale, we tested whether frozen samples from different individuals can be pooled for snRNA-seq, followed by computational demultiplexing, as a cost-effective and scalable approach (*167*), as previously applied to large-scale studies of human peripheral blood mononuclear cells (*24*). We processed lung or prostate samples jointly from three individuals using the CST and TST protocols. We pooled tissue samples from three individuals and processed the pool for single-nucleus extraction, thus minimizing technical batch effects and wet-lab time. After sequencing, we removed ambient RNA (*30*) and performed de novo genotype-based demultiplexing to assign nuclei to donors [using souporcell (*168*); fig. S34; (*28*)]. We validated the demultiplexing by comparing the genotype-based assignments to those from an expression-based multinomial logistic classifier that assigned donor identity to each nucleus profile after training with unpooled samples of the same donors, which showed high concordance

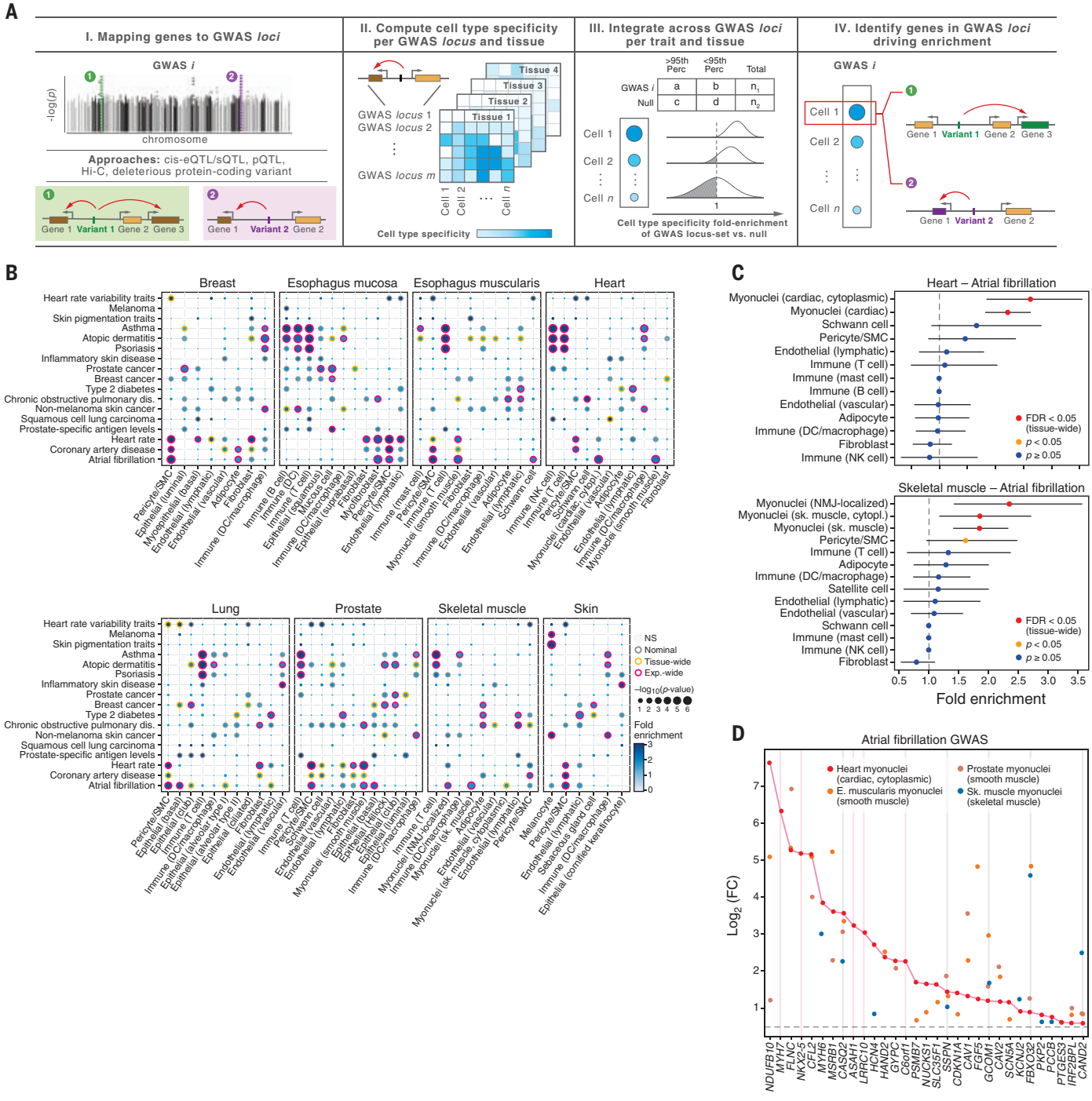


Fig. 6. Cell type–specific enrichment of eQTL and sQTL target genes mapped to GWAS loci. (**A**) Schematic of the method (ECLIPSER). (**B**) Cell-type enrichment of genes mapped to GWAS loci for 17 of the 21 complex traits tested with at least one tissue-wide significant result (FDR < 0.05, correcting for all cell types tested per tissue per trait) across eight GTEx tissues. Gray, orange, and red borders indicate nominal, tissue-wide, and experiment-wide significance (FDR < 0.05, correcting for all cell types tested across eight tissues and 21 traits), respectively. Only cell types with at least one tissue-wide enrichment are shown. (**C** and **D**) Myonuclei and pericyte genes enriched in atrial fibrillation GWAS loci (tissue-wide FDR < 0.05, Bayesian Fisher's exact test). Fold-enrichment (*x* axis) of cell types (*y* axis) for atrial fibrillation GWAS in heart (top) and skeletal muscle (bottom) is shown in (C). Error bars represent 95% credible intervals. Red indicates tissue-wide significance, orange indicates nominal significance, and blue indicates nonsignificance ($P \geq 0.05$, Bayesian Fisher's exact test). Differential expression in myonuclei versus other cell types from heart (red), skeletal muscle (blue), esophagus muscularis (orange), and prostate (brown) of the genes (*x* axis) driving enrichment of atrial fibrillation GWAS loci in heart cardiac myonuclei is shown in (D). Gray and pink vertical lines indicate $\log_2$(fold change) > 0.5 and FDR < 0.1 in myonuclei in all four tissues or only in heart, respectively. FC, fold change.

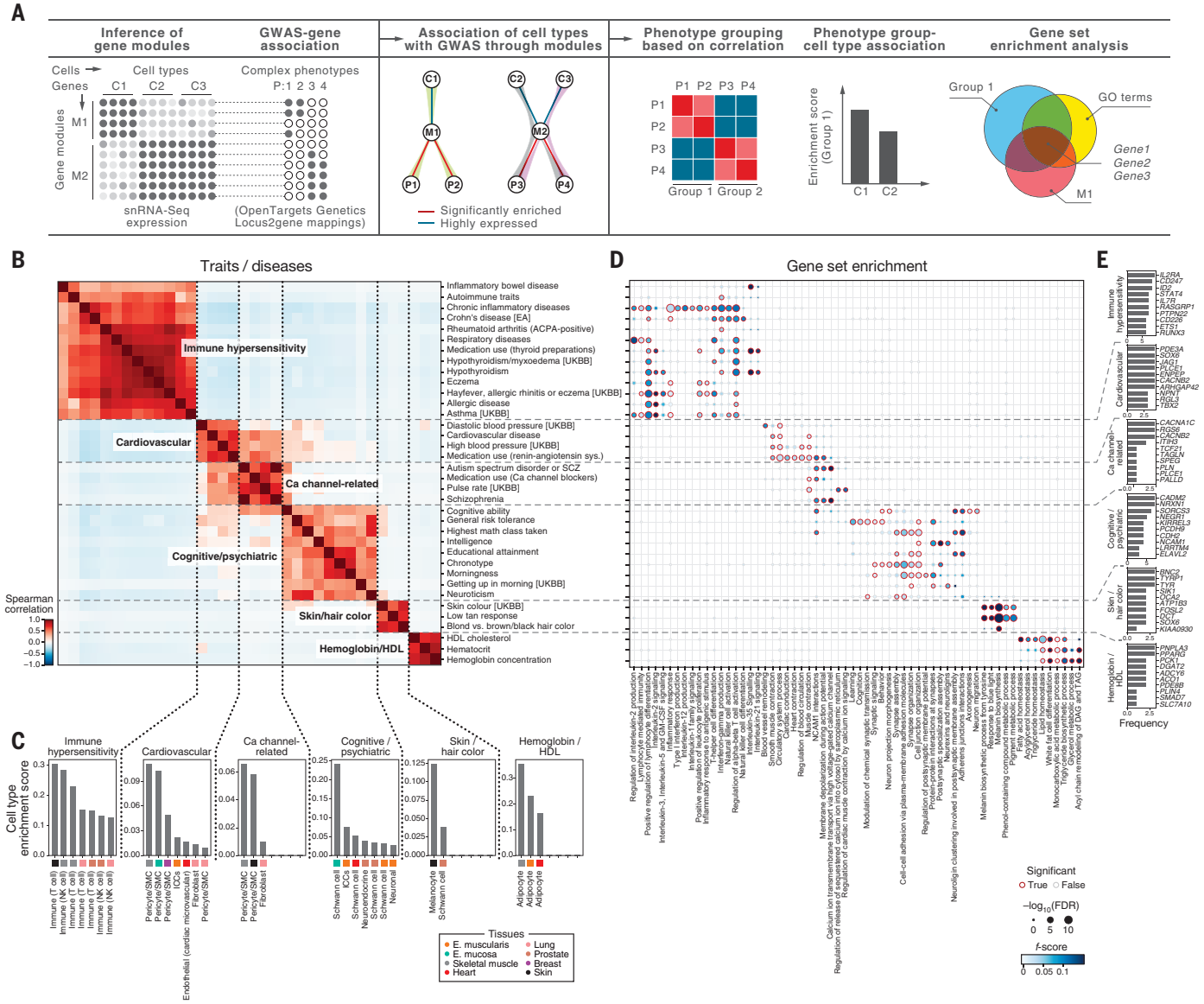


Fig. 7. Cell types and gene modules relevant for trait and disease groups by GWAS module enrichment. (**A**) Schematic of the module-based enrichment method. Shaded edges indicate associations between cell types and phenotypes through modules (middle). (**B** to **E**) Trait and disease groups identified by GWAS-cell type relationships. Similarity (Spearman correlation coefficient) between GWAS traits and diseases (rows and columns) by enriched cell types is shown in (B). Dashed lines demarcate trait and disease groups. Shown in (C) is the cell-type enrichment for each of the GWAS-enriched modules. Significance [$-\log_{10}$(FDR), circle size] and *F* score (circle color) of enrichment of gene sets (columns) with the genes in the intersection of a gene module and GWAS genes for each trait or disease in (B) (rows) are shown in (D). A red outline indicates an FDR less than 0.1. Shown in (E) is the number of traits (*x* axis) in each module in (B) where a gene (*y* axis) is detected as the driver of the association in the intersection of the gene modules, a trait or disease enriched in the module, and a functional gene set for the top 10 most frequently identified genes in the enrichment analysis of each module. EA, East Asian; SCZ, Schizophrenia; UKBB, UK Biobank.

between the two approaches (accuracy 88 to 96%; fig. S34). Moreover, genotype-based doublet calls were concordant with expression-based doublet calls in both lung and prostate (mean balanced accuracy of 63%) (fig. S34).

Discussion

Cross-tissue atlases allow us to characterize tissue-specific and tissue-agnostic features of cells of a common type that serve accessory roles in tissues, such as immune and stroma cells (fig. S35A). For example, for *LYVE1*- and *HLA*II–expressing macrophages, our results reinforce the notion of functional specification of these two macrophage states into tissue support and tissue immunity, respectively (*47*), and propose a model for their differentiation (fig. S35, B and C). In mice, *Lyve1*^{high} macrophages are localized perivascularly, whereas MHCII^{high} macrophages are found in proximity to neurons (*47*). Future studies can address this localization in humans, and signals that govern the tissue-specific ratios of *LYVE1*- versus *HLA*II–expressing cells.

Our data demonstrate the prevalence of LAM-like cells across tissue contexts and pathologies, including breast and heart, where we recovered both adipocytes and LAM-like cells (Figs. 1 and 3C and fig. S2). In line with a model of lipid-induced differentiation of macrophages toward the LAM state (*85*), our classifier recovered LAMs in pathologies characterized by lipid accumulation: atherosclerosis, Crohn's disease, and acne. The inferred role for *PPARG* and *NR1H3* in driving the LAM expression program suggests a model in which signaling

through lipid-bound receptors on macrophages, such as *TREM2* or *CD36*, up-regulates the expression of more lipid receptors, as well as of lipid-modifying enzymes through *PPARG* and *NR1H3*. Because we observed LAMs in healthy organs and in conditions linked to lipid accumulation, future studies may identify other tissue-specific signals or conditions that can trigger LAM-like states.

In lung alveolar fibroblasts, we identified an interconnected functional module that may enable an adequate response to alveolar distortion, through transduction of mechanical cues from the ECM to cytosolic calcium-induced cytoskeletal contraction (fig. S35, D and E). These include BM adhesion genes, including *ITGA8* and its interacting partner *NPNT* (*95*–*100*); a calcium transport module featuring neurotransmitter receptors and multiple calcium ion channels, including the mechanosensitive channel *PIEZO2*; and an actomyosin contractility program that features *MYLK* and mediates cytosolic calcium-induced cytoskeletal contraction, which is required for the transduction of mechanical cues from the ECM (*169*). *PIEZO2* and *ITGA8* are also coexpressed in intraglomerular mesangial cells (*170*, *171*), where *PIEZO2* may sense mechanical forces resulting from changes in blood flow and *ITGA8* confers contractility and adhesion (*172*). Because the alveolus and glomerulus are stretch sensors of air pressure and blood pressure, respectively, *PIEZO2*$^+$/*ITGA8*$^+$ alveolar fibroblasts may contribute to mechanosensing of alveolar tension, which so far has been mainly attributed to Piezo2$^+$ sensory neurons in mice (*111*), suggesting conserved features in mechanosensing across organs. Both lung alveolar fibroblasts and myofibroblasts have been localized to alveoli (*40*), suggesting a possible relationship, possibly through *FGFR4* in alveolar fibroblasts and its ligand *FGF18* in myofibroblasts (fig. S18D) (*109*, *173*).

We further demonstrated the utility of our tissue atlas for monogenic and polygenic disease biology. Many monogenic disease gene modules that are defined by comorbidity [e.g., diabetes and lipodystrophy (*174*)] or similarity of clinical phenotypes (e.g., muscle diseases) were enriched in expected cell types (fig. S35F). Focusing on the pathobiology of monogenic muscle diseases (fig. S35G), we highlighted nonmyocyte cell populations with a potential role in muscle diseases, including nervous system, immune, and stromal cells (*118*), as well as specific myonuclei subsets that express muscle disease genes. Whether the multiple myonuclei subsets we observed are related to multinucleation and specialization of different nuclei in one syncytium (*31*) remains unclear. Some disease-risk genes may also disrupt cell-cell interactions in the muscle. Notably, we observed varying levels of *DMD* expression across cell types, as well as between human and mouse. Variation in *DMD* isoform expression, which has critical implications in Duchenne muscular dystrophy (*135*), can be investigated in future studies. For common complex diseases, we found significant enrichment in specific cell groups for multiple traits (fig. S35H). For more than half of the traits, there was enrichment for the same cell type in different tissues driven by both common and tissue-specific genes. Future work will be needed to extend these analyses across a broad set of tissues and cell types and examine the role of disease-associated genes and cell types in a disease context in patient samples.

Advances in single-cell epigenomics (*175*) and multi-omics (*176*–*178*) should further enable linking GWAS variants to their target genes and the cell types and programs in which they act. Recent findings indicating that a large fraction of genetic regulatory effects linked to GWAS variants can only be detected at the cellular level (*11*, *25*) suggest that cell-level eQTL maps will be essential. The experimental and computational methods we developed for a cross-tissue atlas, and the biological queries we defined, will provide a basis for scaling such efforts to hundreds of individuals and diverse populations.

Methods summary

Tissue samples were selected from among a subset of GTEx project samples that were flash frozen and banked. Nuclei from each sample were extracted using the EZ, CST, NST, and TST protocols described in (*26*). Libraries for snRNA-seq were generated using the Chromium Single Cell 3′ v2 Reagent Kit (10x Genomics), and sequencing was performed with Illumina HiSeq X (96 samples) or NextSeq (three samples), according to the manufacturer's protocols. The resulting snRNA-seq data were aligned and quantified using CellRanger v2.1.0 (10x Genomics), ambient RNA correction was performed using CellBender (*30*), and low-quality nuclei were filtered out using standard criteria (*28*). The resulting snRNA-seq expression profiles were integrated across samples using a total correlation variational autoencoder (*28*). Detailed descriptions of all computational analyses are provided in (*28*).

REFERENCES AND NOTES

1. V. Tam *et al*., Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019). doi: 10.1038/s41576-019-0127-1; pmid: 31068683
2. M. C. Mills, C. Rahal, A scientometric review of genome-wide association studies. *Commun. Biol.* **2**, 9 (2019). doi: 10.1038/s42003-018-0261-x; pmid: 30623105
3. E. Cano-Gamez, G. Trynka, From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* **11**, 424 (2020). doi: 10.3389/fgene.2020.00424; pmid: 32477401
4. J. G. Camp, R. Platt, B. Treutlein, Mapping human cell phenotypes to genotypes with single-cell genomics. *Science* **365**, 1401–1405 (2019). doi: 10.1126/science.aax6648; pmid: 31604266
5. G. Sun *et al*., Single-cell RNA sequencing in cancer: Applications, advances, and emerging challenges. *Mol. Ther. Oncolytics* **21**, 183–206 (2021). doi: 10.1016/j.omto.2021.04.001; pmid: 34027052
6. S. S. Potter, Single-cell RNA sequencing for the study of development, physiology and disease. *Nat. Rev. Nephrol.* **14**, 479–492 (2018). doi: 10.1038/s41581-018-0021-7; pmid: 29789704
7. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). doi: 10.1126/science.aaz1776; pmid: 32913098
8. S. Kasela *et al*., Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4$^+$ versus CD8$^+$ T cells. *PLOS Genet.* **13**, e1006643 (2017). doi: 10.1371/journal.pgen.1006643; pmid: 28248954
9. M. G. P. van der Wijst *et al*., Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018). doi: 10.1038/s41588-018-0089-9; pmid: 29610479
10. D. V. Zhernakova *et al*., Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017). doi: 10.1038/ng.3737; pmid: 27918533
11. S. Kim-Hellmuth *et al*., Cell type-specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020). doi: 10.1126/science.aaz8528; pmid: 32913075
12. C. J. Ye *et al*., Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345**, 1254665 (2014). doi: 10.1126/science.1254665; pmid: 25214635
13. A. S. E. Cuomo *et al*., Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* **11**, 810 (2020). doi: 10.1038/s41467-020-14457-z; pmid: 32041960
14. B. J. Strober *et al*., Dynamic genetic regulation of gene expression during cellular differentiation. *Science* **364**, 1287–1290 (2019). doi: 10.1126/science.aaw0040; pmid: 31249060
15. N. G. Skene *et al*., Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018). doi: 10.1038/s41588-018-0129-5; pmid: 29785013
16. C. S. Smillie *et al*., Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell* **178**, 714–730.e22 (2019). doi: 10.1016/j.cell.2019.06.029; pmid: 31348891
17. E. M. Weeks *et al*., Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. medRxiv 2020.09.08.20190561 [Preprint] (2020); https://doi.org/10.1101/2020.09.08.20190561.
18. K. A. Jagadeesh *et al*., Identifying disease-critical cell types and cellular processes across the human body by integration of single-cell profiles and human genetics. bioRxiv 2021.03.19.436212 [Preprint] (2021); https://doi.org/10.1101/2021.03.19.436212.
19. N. Habib *et al*., Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017). doi: 10.1038/nmeth.4407; pmid: 28846088
20. N. Habib *et al*., Div-Seq: Single-nucleus RNA-seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925–928 (2016). doi: 10.1126/science.aad7038; pmid: 27471252
21. M. J. Petrany *et al*., Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat. Commun.* **11**, 6374 (2020). doi: 10.1038/s41467-020-20063-w; pmid: 33311464
22. H. Wu, Y. Kirita, E. L. Donnelly, B. D. Humphreys, Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: Rare cell types and novel cell states revealed in fibrosis. *J. Am. Soc. Nephrol.* **30**, 23–32 (2019). doi: 10.1681/ASN.2018090912; pmid: 30510133
23. W. Sun *et al*., snRNA-seq reveals a subpopulation of adipocytes that regulates thermogenesis. *Nature* **587**, 98–102 (2020). doi: 10.1038/s41586-020-2856-x; pmid: 33116305
24. H. M. Kang *et al*., Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018). doi: 10.1038/nbt.4042; pmid: 29227470
25. M. van der Wijst *et al*., The single-cell eQTLGen consortium. *eLife* **9**, e52155 (2020). doi: 10.7554/eLife.52155; pmid: 32149610
26. E. Drokhlyansky *et al*., The human and mouse enteric nervous system at single-cell resolution. *Cell* **182**, 1606–1622.e23 (2020). doi: 10.1016/j.cell.2020.08.003; pmid: 32888429
27. T. M. Delorey *et al*., COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* **595**, 107–113 (2021). doi: 10.1038/s41586-021-03570-8; pmid: 33915569

28. Materials and methods are available as supplementary materials.
29. M. Slyper *et al*., A single-cell and single-nucleus RNA-seq toolbox for fresh and frozen human tumors. *Nat. Med.* **26**, 792–802 (2020). doi: 10.1038/s41591-020-0844-1; pmid: 32405060
30. S. J. Fleming, J. C. Marioni, M. Babadi, CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. bioRxiv 791699 [Preprint] (2019); https://doi.org/10.1101/791699.
31. M. Kim *et al*., Single-nucleus transcriptomics reveals functional compartmentalization in syncytial skeletal muscle cells. *Nat. Commun.* **11**, 6375 (2020). doi: 10.1038/s41467-020-20064-9; pmid: 33311457
32. M. Wolfien *et al*., Single-nucleus sequencing of an entire mammalian heart: Cell type composition and velocity. *Cells* **9**, 318 (2020). doi: 10.3390/cells9020318; pmid: 32013057
33. N. R. Tucker *et al*., Transcriptional and cellular diversity of the human heart. *Circulation* **142**, 466–482 (2020). doi: 10.1161/CIRCULATIONAHA.119.045401; pmid: 32403949
34. M. Litviňuková *et al*., Cells of the adult human heart. *Nature* **588**, 466–472 (2020). doi: 10.1038/s41586-020-2797-4; pmid: 32971526
35. R. X. Verma *et al*., Single cell RNA-seq analysis of the flexor digitorum brevis mouse myofibers. *Skelet. Muscle* **11**, 13 (2021). doi: 10.1186/s13395-021-00269-2; pmid: 34001262
36. A. B. Rubenstein *et al*., Single-cell transcriptional profiles in human skeletal muscle. *Sci. Rep.* **10**, 229 (2020). doi: 10.1038/s41598-019-57110-6; pmid: 31937892
37. K. R. Jessen, R. Mirsky, The success and failure of the Schwann cell response to nerve injury. *Front. Cell. Neurosci.* **13**, 33 (2019). doi: 10.3389/fncel.2019.00033; pmid: 30804758
38. G. H. Henry *et al*., A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell Rep.* **25**, 3530–3542.e5 (2018). doi: 10.1016/j.celrep.2018.11.086; pmid: 30566875
39. M. D. Luecken, F. J. Theis, Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019). doi: 10.15252/msb.20188746; pmid: 31217225
40. K. J. Travaglini *et al*., A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020). doi: 10.1038/s41586-020-2922-4; pmid: 33208946
41. S. C. van den Brink *et al*., Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017). doi: 10.1038/nmeth.4437; pmid: 28960196
42. E. Denisenko *et al*., Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* **21**, 130 (2020). doi: 10.1186/s13059-020-02048-6; pmid: 32487174
43. A. Zaghlool *et al*., Characterization of the nuclear and cytosolic transcriptomes in human brain tissue reveals new insights into the subcellular distribution of RNA transcripts. *Sci. Rep.* **11**, 4076 (2021). doi: 10.1038/s41598-021-83541-1; pmid: 33603054
44. M. Rabani *et al*., High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698–1710 (2014). doi: 10.1016/j.cell.2014.11.015; pmid: 25497548
45. B. W. Solnestam *et al*., Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs. *BMC Genomics* **13**, 574 (2012). doi: 10.1186/1471-2164-13-574; pmid: 23110385
46. W. L. Hwang *et al*., Single-nucleus and spatial transcriptomics of archival pancreatic cancer reveals multi-compartment reprogramming after neoadjuvant treatment. bioRxiv 2020.08.25.267336 [Preprint] (2020); https://doi.org/10.1101/2020.08.25.267336.
47. S. Chakarov *et al*., Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches. *Science* **363**, eaau0964 (2019). doi: 10.1126/science.aau0964; pmid: 30872492
48. C. Varol, A. Mildner, S. Jung, Macrophages: Development and tissue specialization. *Annu. Rev. Immunol.* **33**, 643–675 (2015). doi: 10.1146/annurev-immunol-032414-112220; pmid: 25861979
49. A.-C. Villani *et al*., Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017). doi: 10.1126/science.aah4573; pmid: 28428369
50. P. A. Reyfman *et al*., Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019). doi: 10.1164/rccm.201712-2410OC; pmid: 30554520
51. D. A. Jaitin *et al*., Lipid-associated macrophages control metabolic homeostasis in a Trem2-dependent manner. *Cell* **178**, 686–698.e14 (2019). doi: 10.1016/j.cell.2019.05.054; pmid: 31257031
52. T. Capucha *et al*., Distinct murine mucosal langerhans cell subsets develop from pre-dendritic cells and monocytes. *Immunity* **43**, 369–381 (2015). doi: 10.1016/j.immuni.2015.06.017; pmid: 26231115
53. J. Deckers, H. Hammad, E. Hoste, Langerhans cells: Sensing the environment in health and disease. *Front. Immunol.* **9**, 93 (2018). doi: 10.3389/fimmu.2018.00093; pmid: 29449841
54. E. Evren *et al*., Distinct developmental pathways from blood monocytes generate human lung macrophage diversity. *Immunity* **54**, 259–275.e7 (2021). doi: 10.1016/j.immuni.2020.12.003; pmid: 33382972
55. A. A. Patel *et al*., The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* **214**, 1913–1923 (2017). doi: 10.1084/jem.20170355; pmid: 28606987
56. H. Y. Lim *et al*., Hyaluronan receptor LYVE-1-expressing macrophages maintain arterial tone through hyaluronan-mediated regulation of smooth muscle cell collagen. *Immunity* **49**, 326–341.e7 (2018). doi: 10.1016/j.immuni.2018.06.008; pmid: 30054204
57. M. Bizou *et al*., Cardiac macrophage subsets differentially regulate lymphatic network remodeling during pressure overload. *Sci. Rep.* **11**, 16801 (2021). doi: 10.1038/s41598-021-95723-y; pmid: 34413352
58. P. Ramachandran *et al*., Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* **575**, 512–518 (2019). doi: 10.1038/s41586-019-1631-3; pmid: 31597160
59. M. J. Perugorria *et al*., Non-parenchymal TREM-2 protects the liver from immune-mediated hepatocellular damage. *Gut* **68**, 533–546 (2019). doi: 10.1136/gutjnl-2017-314107; pmid: 29374630
60. X. Xiong *et al*., Landscape of intercellular crosstalk in healthy and NASH liver revealed by single-cell secretome gene analysis. *Mol. Cell* **75**, 644–660.e5 (2019). doi: 10.1016/j.molcel.2019.07.028; pmid: 31398325
61. M. Guilliams *et al*., Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379–396.e38 (2022). doi: 10.1016/j.cell.2021.12.018; pmid: 35021063
62. E. A. Ayaub *et al*., Single cell RNA-seq and mass cytometry reveals a novel and a targetable population of macrophages in idiopathic pulmonary fibrosis. bioRxiv 2021.01.04.425268 [Preprint] (2021); https://doi.org/10.1101/2021.01.04.425268.
63. C. Cochain *et al*., Single-cell RNA-seq reveals the transcriptional landscape and heterogeneity of aortic macrophages in murine atherosclerosis. *Circ. Res.* **122**, 1661–1674 (2018). doi: 10.1161/CIRCRESAHA.117.312509; pmid: 29545365
64. L. Willemsen, M. P. J. de Winther, Macrophage subsets in atherosclerosis as defined by single-cell technologies. *J. Pathol.* **250**, 705–714 (2020). doi: 10.1002/path.5392; pmid: 32003464
65. F. Ma *et al*., The cellular architecture of the antimicrobial response network in human leprosy granulomas. *Nat. Immunol.* **22**, 839–850 (2021). doi: 10.1038/s41590-021-00956-8; pmid: 34168371
66. H. Keren-Shaul *et al*., A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017). doi: 10.1016/j.cell.2017.05.018; pmid: 28602351
67. A. Grubman *et al*., A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* **22**, 2087–2097 (2019). doi: 10.1038/s41593-019-0539-4; pmid: 31768052
68. H. Mathys *et al*., Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019). doi: 10.1038/s41586-019-1195-2; pmid: 31042697
69. C. Muus *et al*., Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat. Med.* **27**, 546–559 (2021). doi: 10.1038/s41591-020-01227-z; pmid: 33654293
70. R. C. Wirka *et al*., Atheroprotective roles of smooth muscle cell phenotypic modulation and the *TCF21* disease gene as revealed by single-cell analysis. *Nat. Med.* **25**, 1280–1289 (2019). doi: 10.1038/s41591-019-0512-5; pmid: 31359001
71. T. Alsaigh, D. Evans, D. Frankel, A. Torkamani, Decoding the transcriptome of atherosclerotic plaque at single-cell resolution. bioRxiv 2020.03.03.968123 [Preprint] (2020); https://doi.org/10.1101/2020.03.03.968123.
72. R. Vento-Tormo *et al*., Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018). doi: 10.1038/s41586-018-0698-6; pmid: 30429548
73. K. Tan *et al*., Transcriptome profiling reveals signaling conditions dictating human spermatogonia fate in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17832–17841 (2020). doi: 10.1073/pnas.2000362117; pmid: 32661178
74. J. Liao *et al*., Single-cell RNA sequencing of human kidney. *Sci. Data* **7**, 4 (2020). doi: 10.1038/s41597-019-0351-8; pmid: 31896769
75. M. M. F. Qadir *et al*., Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10876–10887 (2020). doi: 10.1073/pnas.1918314117; pmid: 32354994
76. S. A. MacParland *et al*., Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018). doi: 10.1038/s41467-018-06318-7; pmid: 30348985
77. T. Chitiashvili *et al*., Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. *Nat. Cell Biol.* **22**, 1436–1446 (2020). doi: 10.1038/s41556-020-00607-4; pmid: 33257808
78. A. J. De Micheli, J. A. Spector, O. Elemento, B. D. Cosgrove, A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations. *Skelet. Muscle* **10**, 19 (2020). doi: 10.1186/s13395-020-00236-3; pmid: 32624006
79. T. K. Hughes *et al*., Second-strand synthesis-based massively parallel scRNA-Seq reveals cellular states and molecular features of human inflammatory skin pathologies. *Immunity* **53**, 878–894.e7 (2020). doi: 10.1016/j.immuni.2020.09.015; pmid: 33053333
80. T. B. Rojahn *et al*., Single-cell transcriptomics combined with interstitial fluid proteomics defines cell type-specific immune regulation in atopic dermatitis. *J. Allergy Clin. Immunol.* **146**, 1056–1069 (2020). doi: 10.1016/j.jaci.2020.03.041; pmid: 32344053
81. K. Rindler *et al*., Spontaneously resolved atopic dermatitis shows melanocyte and immune cell activation distinct from healthy control skin. *Front. Immunol.* **12**, 630892 (2021). doi: 10.3389/fimmu.2021.630892; pmid: 33717163
82. C. W. Y. Ha *et al*., Translocation of viable gut microbiota to mesenteric adipose drives formation of creeping fat in humans. *Cell* **183**, 666–683.e17 (2020). doi: 10.1016/j.cell.2020.09.009; pmid: 32991841
83. J. Tome-Garcia *et al*., Cell type-specific isolation and transcriptomic profiling informs glial pathology in human temporal lobe epilepsy. bioRxiv 2020.12.11.421370 [Preprint] (2020); https://doi.org/10.1101/2020.12.11.421370.
84. A. Iyer *et al*., Increased chitotriosidase activity in serum of leprosy patients: Association with bacillary leprosy. *Clin. Immunol.* **131**, 501–509 (2009). doi: 10.1016/j.clim.2009.02.003; pmid: 19307157
85. A. Deczkowska, A. Weiner, I. Amit, The physiology, pathology, and potential therapeutic applications of the TREM2 signaling pathway. *Cell* **181**, 1207–1217 (2020). doi: 10.1016/j.cell.2020.05.003; pmid: 32531244
86. Y. Kidani, S. J. Bensinger, Liver X receptor and peroxisome proliferator-activated receptor as integrators of lipid homeostasis and immunity. *Immunol. Rev.* **249**, 72–83 (2012). doi: 10.1111/j.1600-065X.2012.01153.x; pmid: 22889216
87. J. C. Savage *et al*., Nuclear receptors license phagocytosis by Trem2^{+} myeloid cells in mouse models of Alzheimer's disease. *J. Neurosci.* **35**, 6532–6543 (2015). doi: 10.1523/JNEUROSCI.4586-14.2015; pmid: 25904803
88. M. V. Plikus *et al*., Fibroblasts: Origins, definitions, and functions in health and disease. *Cell* **184**, 3852–3872 (2021). doi: 10.1016/j.cell.2021.06.024; pmid: 34297930
89. S. Pikkarainen, H. Tokola, R. Kerkelä, H. Ruskoaho, GATA transcription factors in the developing and adult heart. *Cardiovasc. Res.* **63**, 196–207 (2004). doi: 10.1016/j.cardiores.2004.03.025; pmid: 15249177
90. G. M. Dittrich *et al*., Fibroblast GATA-4 and GATA-6 promote myocardial adaptation to pressure overload by enhancing cardiac angiogenesis. *Basic Res. Cardiol.* **116**, 26 (2021). doi: 10.1007/s00395-021-00862-y; pmid: 33876316
91. E. Forte *et al*., Adult fibroblasts retain organ-specific transcriptomic identity. bioRxiv 2021.06.03.446915 [Preprint] (2021); https://doi.org/10.1101/2021.06.03.446915.
92. M. B. Furtado *et al*., Cardiogenic genes expressed in cardiac fibroblasts contribute to heart development and repair. *Circ. Res.* **114**, 1422–1434 (2014). doi: 10.1161/CIRCRESAHA.114.302530; pmid: 24650916

93. S. J. Forrester *et al*., Angiotensin II signal transduction: An update on mechanisms of physiology and pathophysiology. *Physiol. Rev.* **98**, 1627–1738 (2018). doi: 10.1152/physrev.00038.2017; pmid: 29873596
94. S. K. Powers, A. B. Morton, H. Hyatt, M. J. Hinkley, The renin-angiotensin system and skeletal muscle. *Exerc. Sport Sci. Rev.* **46**, 205–214 (2018). doi: 10.1249/JES.0000000000000158; pmid: 30001274
95. D. Kiyozumi *et al*., Basement membrane assembly of the integrin α8β1 ligand nephronectin requires Fraser syndrome-associated proteins. *J. Cell Biol.* **197**, 677–689 (2012). doi: 10.1083/jcb.201203065; pmid: 22613833
96. D. Kiyozumi, N. Sugimoto, K. Sekiguchi, Breakdown of the reciprocal stabilization of QBRICK/Frem1, Fras1, and Frem2 at the basement membrane provokes Fraser syndrome-like defects. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11981–11986 (2006). doi: 10.1073/pnas.0601011103; pmid: 16880404
97. R. Brandenberger *et al*., Identification and characterization of a novel extracellular matrix protein nephronectin that is associated with integrin α8β1 in the embryonic kidney. *J. Cell Biol.* **154**, 447–458 (2001). doi: 10.1083/jcb.200103069; pmid: 11470831
98. Y. Sato *et al*., Molecular basis of the recognition of nephronectin by integrin α8β1. *J. Biol. Chem.* **284**, 14524–14536 (2009). doi: 10.1074/jbc.M900200200; pmid: 19342381
99. U. Müller *et al*., Integrin α8β1 is critically important for epithelial-mesenchymal interactions during kidney morphogenesis. *Cell* **88**, 603–613 (1997). doi: 10.1016/S0092-8674(00)81903-0; pmid: 9054500
100. J. T. Benjamin *et al*., The role of integrin α8β1 in fetal lung morphogenesis and injury. *Dev. Biol.* **335**, 407–417 (2009). doi: 10.1016/j.ydbio.2009.09.021; pmid: 19769957
101. S. Vrontou *et al*., *Fras1* deficiency results in cryptophthalmos, renal agenesis and blebbed phenotype in mice. *Nat. Genet.* **34**, 209–214 (2003). doi: 10.1038/ng1168; pmid: 12766770
102. P. Petrou, E. Pavlakis, Y. Dalezios, V. K. Galanopoulos, G. Chalepakis, Basement membrane distortions impair lung lobation and capillary organization in the mouse model for Fraser syndrome. *J. Biol. Chem.* **280**, 10350–10356 (2005). doi: 10.1074/jbc.M412368200; pmid: 15623520
103. J. M. Linton, G. R. Martin, L. F. Reichardt, The ECM protein nephronectin promotes kidney development via integrin α8β1-mediated stimulation of *Gdnf* expression. *Development* **134**, 2501–2509 (2007). doi: 10.1242/dev.005033; pmid: 17537792
104. L. McGregor *et al*., Fraser syndrome and mouse blebbed phenotype caused by mutations in *FRAS1/Fras1* encoding a putative extracellular matrix protein. *Nat. Genet.* **34**, 203–208 (2003). doi: 10.1038/ng1142; pmid: 12766769
105. Y. Shafeghati, A. Kniepert, G. Vakili, M. Zenker, Fraser syndrome due to homozygosity for a splice site mutation of *FREM2*. *Am. J. Med. Genet. A.* **146A**, 529–531 (2008). doi: 10.1002/ajmg.a.32091; pmid: 18203166
106. A. M. Alazami *et al*., *FREM1* mutations cause bifid nose, renal agenesis, and anorectal malformations syndrome. *Am. J. Hum. Genet.* **85**, 414–418 (2009). doi: 10.1016/j.ajhg.2009.08.010; pmid: 19732862
107. T. J. Carney *et al*., Genetic analysis of fin development in zebrafish identifies Furin and Hemicentin1 as potential novel Fraser syndrome disease genes. *PLOS Genet.* **6**, e1000907 (2010). doi: 10.1371/journal.pgen.1000907; pmid: 20419147
108. M. B. Buechler *et al*., Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021). doi: 10.1038/s41586-021-03549-5; pmid: 33981032
109. M. Weinstein, X. Xu, K. Ohyama, C. X. Deng, FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development* **125**, 3615–3623 (1998). doi: 10.1242/dev.125.18.3615; pmid: 9716527
110. M. Rezvani *et al*., Association of a FGFR-4 gene polymorphism with bronchopulmonary dysplasia and neonatal respiratory distress. *Dis. Markers* **35**, 633–640 (2013). doi: 10.1155/2013/932356; pmid: 24288432
111. K. Nonomura *et al*., Piezo2 senses airway stretch and mediates lung inflation-induced apnoea. *Nature* **541**, 176–181 (2017). doi: 10.1038/nature20793; pmid: 28002412
112. A. Delle Vedove *et al*., Biallelic loss of proprioception-related PIEZO2 causes muscular atrophy with perinatal respiratory distress, arthrogryposis, and scoliosis. *Am. J. Hum. Genet.* **99**, 1206–1216 (2016). doi: 10.1016/j.ajhg.2016.09.019; pmid: 27843126
113. I. Y. Kuo, B. E. Ehrlich, Signaling in muscle contraction. *Cold Spring Harb. Perspect. Biol.* **7**, a006023 (2015). doi: 10.1101/cshperspect.a006023; pmid: 25646377
114. D. T. Montoro *et al*., A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018). doi: 10.1038/s41586-018-0393-7; pmid: 30069044
115. L. W. Plasschaert *et al*., A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381 (2018). doi: 10.1038/s41586-018-0394-6; pmid: 30069046
116. J. S. Amberger, C. A. Bocchini, A. F. Scott, A. Hamosh, OMIM.org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* **47**, D1038–D1043 (2019). doi: 10.1093/nar/gky1151; pmid: 30445645
117. K. Inaba, K. Mizuno, Sperm dysfunction and ciliopathy. *Reprod. Med. Biol.* **15**, 77–94 (2015). doi: 10.1007/s12522-015-0225-5; pmid: 29259424
118. L. Benarroch, G. Bonne, F. Rivier, D. Hamroun, The 2021 version of the gene table of neuromuscular disorders (nuclear genome). *Neuromuscul. Disord.* **30**, 1008–1048 (2020). doi: 10.1016/j.nmd.2020.11.009; pmid: 33257164
119. R. Csapo, M. Gumpenberger, B. Wessner, Skeletal muscle extracellular matrix – What do we know about its composition, regulation, and physiological roles? A narrative review. *Front. Physiol.* **11**, 253 (2020). doi: 10.3389/fphys.2020.00253; pmid: 32265741
120. K. Ohlendieck, D. Swandulla, Complexity of skeletal muscle degeneration: Multi-systems pathophysiology and organ crosstalk in dystrophinopathy. *Pflugers Arch.* **473**, 1813–1839 (2021). doi: 10.1007/s00424-021-02623-1; pmid: 34553265
121. J. Verschuuren, E. Strijbos, A. Vincent, Neuromuscular junction disorders. *Handb. Clin. Neurol.* **133**, 447–466 (2016). doi: 10.1016/B978-0-444-63432-0.00024-4; pmid: 27112691
122. B. B. Cummings *et al*., Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* **9**, eaal5209 (2017). doi: 10.1126/scitranslmed.aal5209; pmid: 28424332
123. H. D. Gonorazky *et al*., Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. *Am. J. Hum. Genet.* **104**, 466–483 (2019). doi: 10.1016/j.ajhg.2019.01.012; pmid: 30827497
124. A. E. Vanhaesebrouck, D. Beeson, The congenital myasthenic syndromes: Expanding genetic and phenotypic spectrums and refining treatment strategies. *Curr. Opin. Neurol.* **32**, 696–703 (2019). doi: 10.1097/WCO.0000000000000736; pmid: 31361628
125. A. M. Maggs, C. Huxley, S. M. Hughes, Nerve-dependent changes in skeletal muscle myosin heavy chain after experimental denervation and cross-reinnervation and in a demyelinating mouse model of Charcot-Marie-Tooth disease type 1A. *Muscle Nerve* **38**, 1572–1584 (2008). doi: 10.1002/mus.21106; pmid: 19016545
126. C. G. Bönnemann, The collagen VI-related myopathies: Muscle meets its matrix. *Nat. Rev. Neurol.* **7**, 379–390 (2011). doi: 10.1038/nrneurol.2011.81; pmid: 21691338
127. A. de Morrée *et al*., Dysferlin regulates cell adhesion in human monocytes. *J. Biol. Chem.* **288**, 14147–14157 (2013). doi: 10.1074/jbc.M112.448589; pmid: 23558685
128. J. Lin *et al*., Transcriptional co-activator PGC-1α drives the formation of slow-twitch muscle fibres. *Nature* **418**, 797–801 (2002). doi: 10.1038/nature00904; pmid: 12181572
129. H. Tajsharghi *et al*., Recessive myosin myopathy with external ophthalmoplegia associated with *MYH2* mutations. *Eur. J. Hum. Genet.* **22**, 801–808 (2014). doi: 10.1038/ejhg.2013.250; pmid: 24193343
130. V. Pisani *et al*., Preferential central nucleation of type 2 myofibers is an invariable feature of myotonic dystrophy type 2. *Muscle Nerve* **38**, 1405–1411 (2008). doi: 10.1002/mus.21122; pmid: 18816606
131. A. Vihola *et al*., Histopathological differences of myotonic dystrophy type 1 (DM1) and PROMM/DM2. *Neurology* **60**, 1854–1857 (2003). doi: 10.1212/01.WNL.0000065898.61358.09; pmid: 12796551
132. L. S. Krivickas, T. Ansved, D. Suh, W. R. Frontera, Contractile properties of single muscle fibers in myotonic dystrophy. *Muscle Nerve* **23**, 529–537 (2000). doi: 10.1002/(SICI)1097-4598(200004)23:4<529::AID-MUS11>3.0.CO;2-Y; pmid: 10716763
133. N. F. Clarke, Congenital fiber-type disproportion. *Semin. Pediatr. Neurol.* **18**, 264–271 (2011). doi: 10.1016/j.spen.2011.10.008; pmid: 22172422
134. J. Strakova *et al*., Integrative effects of dystrophin loss on metabolic function of the mdx mouse. *Sci. Rep.* **8**, 13624 (2018). doi: 10.1038/s41598-018-31753-3; pmid: 30206270
135. D. Duan, N. Goemans, S. Takeda, E. Mercuri, A. Aartsma-Rus, Duchenne muscular dystrophy. *Nat. Rev. Dis. Primers* **7**, 13 (2021). doi: 10.1038/s41572-021-00248-3; pmid: 33602943
136. M. Sander *et al*., Functional muscle ischemia in neuronal nitric oxide synthase-deficient skeletal muscle of children with Duchenne muscular dystrophy. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13818–13823 (2000). doi: 10.1073/pnas.250379497; pmid: 11087833
137. P. A. C. 't Hoen *et al*., Generation and characterization of transgenic mice with the full-length human DMD gene. *J. Biol. Chem.* **283**, 5899–5907 (2008). doi: 10.1074/jbc.M709410200; pmid: 18083704
138. M. W. Hogarth *et al*., Fibroadipogenic progenitors are responsible for muscle loss in limb girdle muscular dystrophy 2B. *Nat. Commun.* **10**, 2430 (2019). doi: 10.1038/s41467-019-10438-z; pmid: 31160583
139. D. R. Lemos *et al*., Nilotinib reduces muscle fibrosis in chronic muscle injury by promoting TNF-mediated apoptosis of fibro/adipogenic progenitors. *Nat. Med.* **21**, 786–794 (2015). doi: 10.1038/nm.3869; pmid: 26053624
140. GTEx Consortium, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017). doi: 10.1038/nature24277; pmid: 29022597
141. X. Unamuno *et al*., Dermatopontin, a novel adipokine promoting adipose tissue extracellular matrix remodelling and inflammation in obesity. *J. Clin. Med.* **9**, 1069 (2020). doi: 10.3390/jcm9041069; pmid: 32283761
142. S.-H. Heo, J.-Y. Cho, ELK3 suppresses angiogenesis by inhibiting the transcriptional activity of ETS-1 on MT1-MMP. *Int. J. Biol. Sci.* **10**, 438–447 (2014). doi: 10.7150/ijbs.8095; pmid: 24719561
143. P. Podkalicka, O. Mucha, J. Dulak, A. Loboda, Targeting angiogenesis in Duchenne muscular dystrophy. *Cell. Mol. Life Sci.* **76**, 1507–1528 (2019). doi: 10.1007/s00018-019-03006-7; pmid: 30770952
144. M. Efremova, M. Vento-Tormo, S. A. Teichmann, R. Vento-Tormo, CellPhoneDB: Inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020). doi: 10.1038/s41596-020-0292-x; pmid: 32103204
145. G. Narkis *et al*., Lethal congenital contractural syndrome type 2 (LCCS2) is caused by a mutation in *ERBB3* (*Her3*), a modulator of the phosphatidylinositol-3-kinase/Akt pathway. *Am. J. Hum. Genet.* **81**, 589–595 (2007). doi: 10.1086/520770; pmid: 17701904
146. O. Ibraghimov-Beskrovnaya *et al*., Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature* **355**, 696–702 (1992). doi: 10.1038/355696a0; pmid: 1741056
147. A. R. Gillies, R. L. Lieber, Structure and function of the skeletal muscle extracellular matrix. *Muscle Nerve* **44**, 318–331 (2011). doi: 10.1002/mus.22094; pmid: 21949456
148. S.-Y. Dai *et al*., Galectin-9 induces maturation of human monocyte-derived dendritic cells. *J. Immunol.* **175**, 2974–2981 (2005). doi: 10.4049/jimmunol.175.5.2974; pmid: 16116184
149. H. K. Finucane *et al*., Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015). doi: 10.1038/ng.3404; pmid: 26414678
150. E. R. Gamazon *et al*., Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **50**, 956–967 (2018). doi: 10.1038/s41588-018-0154-4; pmid: 29955180
151. A. N. Barbeira *et al*., Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* **22**, 49 (2021). doi: 10.1186/s13059-020-02252-4; pmid: 33499903
152. J. M. Rouhana *et al*., ECLIPSER: identifying causal cell types and genes for complex traits through single cell enrichment of e/sQTL-mapped genes in GWAS loci. bioRxiv 2021.11.24.469720 [Preprint] (2021); https://doi.org/10.1101/2021.11.24.469720.
153. M. Ghoussaini *et al*., Open Targets Genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021). doi: 10.1093/nar/gkaa840; pmid: 33045747
154. E. Mountjoy *et al*., An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021). doi: 10.1038/s41588-021-00945-5; pmid: 34711957
155. E. Brakenhielm, K. Alitalo, Cardiac lymphatics in health and disease. *Nat. Rev. Cardiol.* **16**, 56–68 (2019). doi: 10.1038/s41569-018-0087-8; pmid: 30333526
156. F. Paneni, J. A. Beckman, M. A. Creager, F. Cosentino, Diabetes and vascular disease: pathophysiology, clinical consequences, and medical therapy: Part I. *Eur. Heart J.* **34**,

2436–2443 (2013). doi: 10.1093/eurheartj/eht149; pmid: 23641007
157. T. R. Einarson, A. Acs, C. Ludwig, U. H. Panton, Prevalence of cardiovascular disease in type 2 diabetes: A systematic literature review of scientific evidence from across the world in 2007–2017. *Cardiovasc. Diabetol.* **17**, 83 (2018). doi: 10.1186/s12933-018-0728-6; pmid: 29884191
158. C. Kothari, C. Diorio, F. Durocher, The importance of breast adipose tissue in breast cancer. *Int. J. Mol. Sci.* **21**, 5760 (2020). doi: 10.3390/ijms21165760; pmid: 32796696
159. A. B. Kay, Allergy and allergic diseases. First of two parts. *N. Engl. J. Med.* **344**, 30–37 (2001). doi: 10.1056/NEJM200101043440106; pmid: 11136958
160. S. J. Galli, M. Tsai, A. M. Piliponsky, The development of allergic inflammation. *Nature* **454**, 445–454 (2008). doi: 10.1038/nature07204; pmid: 18650915
161. H. Yilmaz, M. Cakmak, B. Ceydilek, C. Demir, A. Aktas, Role of interlekin-35 as a biomarker in patients with newly diagnosed Hashimoto's thyroiditis. *Endocr. Regul.* **50**, 55–61 (2016). doi: 10.1515/enr-2016-0009; pmid: 27560637
162. Y. Li *et al*., The possible role of the novel cytokines IL-35 and IL-37 in inflammatory bowel disease. *Mediators Inflamm.* **2014**, 136329 (2014). doi: 10.1155/2014/136329; pmid: 25214710
163. M.-H. Saeed *et al*., Decreased serum levels of IL-27 and IL-35 in patients with Graves' disease. *Arch. Endocrinol. Metab.* **64**, 521–527 (2021). doi: 10.20945/2359-3997000000227; pmid: 34033291
164. A. Pinggera *et al*., New gain-of-function mutation shows CACNA1D as recurrently mutated gene in autism spectrum disorders and epilepsy. *Hum. Mol. Genet.* **26**, 2923–2932 (2017). doi: 10.1093/hmg/ddx175; pmid: 28472301
165. R. Mysore, G. Liebisch, Y. Zhou, V. M. Olkkonen, P. A. Nidhina Haridas, Angiopoietin-like 8 (Angptl8) controls adipocyte lipolysis and phospholipid composition. *Chem. Phys. Lipids* **207**, 246–252 (2017). doi: 10.1016/j.chemphyslip.2017.05.002; pmid: 28528274
166. A. Yang, E. P. Mottillo, Adipocyte lipolysis: From molecular mechanisms of regulation to disease and therapeutics. *Biochem. J.* **477**, 985–1008 (2020). doi: 10.1042/BCJ20190468; pmid: 32168372
167. J. Xu *et al*., Genotype-free demultiplexing of pooled single-cell RNA-seq. *Genome Biol.* **20**, 290 (2019). doi: 10.1186/s13059-019-1852-7; pmid: 31856883
168. H. Heaton *et al*., Souporcell: Robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020). doi: 10.1038/s41592-020-0820-1; pmid: 32366989
169. J. Z. Kechagia, J. Ivaska, P. Roca-Cusachs, Integrins as biomechanical sensors of the microenvironment. *Nat. Rev. Mol. Cell Biol.* **20**, 457–473 (2019). doi: 10.1038/s41580-019-0134-2; pmid: 31182865
170. B. B. Lake *et al*., A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nat. Commun.* **10**, 2832 (2019). doi: 10.1038/s41467-019-10861-2; pmid: 31249312
171. B. B. Lake *et al*., An atlas of healthy and injured cell states and niches in the human kidney. bioRxiv 2021.07.28.454201 [Preprint] (2021); https://doi.org/10.1101/2021.07.28.454201.
172. I. Marek, K. F. Hilgers, W. Rascher, J. Woelfle, A. Hartner, A role for the alpha-8 integrin chain (itga8) in glomerular homeostasis of the kidney. *Mol. Cell Pediatr.* **7**, 13 (2020). doi: 10.1186/s40348-020-00105-5; pmid: 33000355
173. H. Usui *et al*., Fgf18 is required for embryonic lung alveolar development. *Biochem. Biophys. Res. Commun.* **322**, 887–892 (2004). doi: 10.1016/j.bbrc.2004.07.198; pmid: 15336546
174. J. Zammouri *et al*., Molecular and cellular bases of lipodystrophy syndromes. *Front. Endocrinol.* **12**, 803189 (2022). doi: 10.3389/fendo.2021.803189; pmid: 35046902
175. G. Kelsey, O. Stegle, W. Reik, Single-cell epigenomics: Recording the past and predicting the future. *Science* **358**, 69–75 (2017). doi: 10.1126/science.aan6826; pmid: 28983045
176. S. Ma *et al*., Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116.e20 (2020). doi: 10.1016/j.cell.2020.09.056; pmid: 33098772
177. E. P. Mimitou *et al*., Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat. Biotechnol.* **39**, 1246–1258 (2021). doi: 10.1038/s41587-021-00927-2; pmid: 34083792
178. E. Fiskin *et al*., Single-cell profiling of proteins and chromatin accessibility using PHAGE-ATAC. *Nat. Biotechnol.* **40**, 374–381 (2022). doi: 10.1038/s41587-021-01065-5; pmid: 34675424
179. G. Eraslan, The code used in Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. Zenodo (2022); https://doi.org/10.5281/zenodo.5775379.

ACKNOWLEDGMENTS

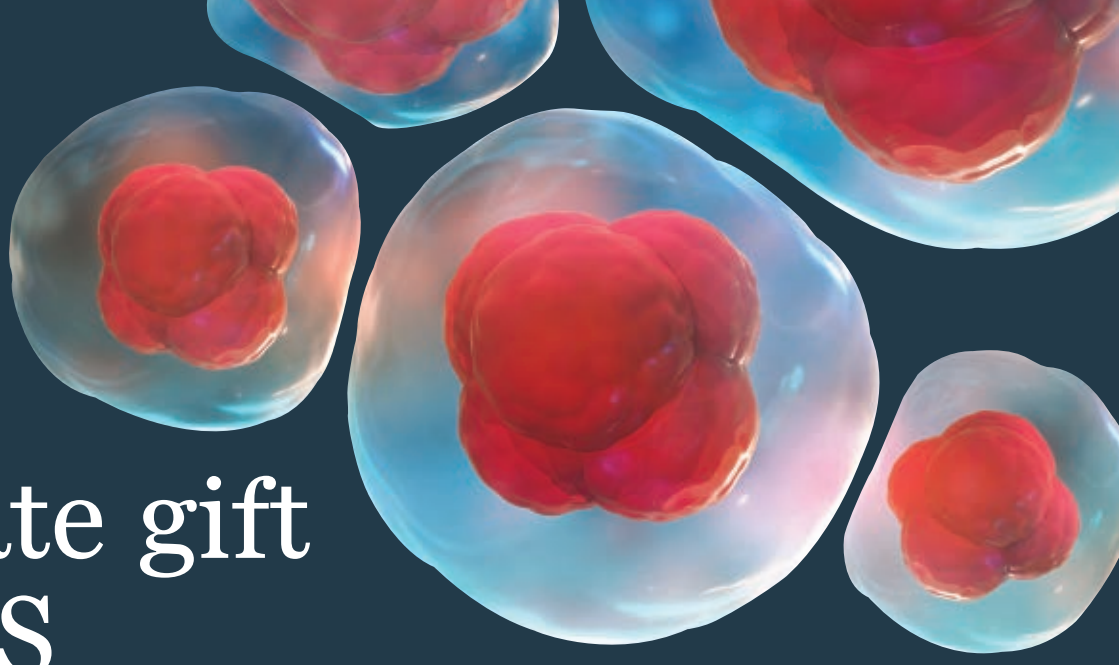
We thank L. Gaffney and A. Hupalowska for help with figure preparation. We thank T. Harvey, B. Cummings, B. Weisburd, A. O. Luria, and D. G. MacArthur for sharing expertise on muscle diseases. We thank A. R. Hamel for assistance with running GeneEnrich. This publication is part of the Human Cell Atlas (www.humancellatlas.org/publications). **Funding:** This project has been funded in part with funds from the Manton Foundation, Klarman Family Foundation, and Howard Hughes Medical Institute (A.R.). This work was also supported in part by the Common Fund of the Office of the Director, US National Institutes of Health (NIH), and by the National Cancer Institute (NCI), National Human Genome Research Institute (NHGRI), National Heart, Lung, and Blood Institute (NHLBI), National Institute on Drug Abuse (NIDA), National Institute of Mental Health (NIMH), National Institute on Aging (NIA), National Institute of Allergy and Infectious Diseases (NIAID), and National Institute of Neurological Disorders and Stroke (NINDS) through NIH contract HHSN268201000029C (S.A., G.G., F.A., and K.G.A.); 5U41HG009494 (S.A., G.G., F.A., and K.G.A.); NEI R01 EY031424-01 (A.V.S. and J.M.R.), and Chan Zuckerberg Initiative (CZI) Seed Network for the Human Cell Atlas award CZF2019-002459 (A.V.S. and J.M.R.). A.M.T. is funded by NHLBI grants R56HL157632 and R21HL156124. **Author contributions:** Conceptualization: K.G.A., O.R.-R., A.R.; Data curation: A.S., D.D., F.A., G.E., O.A., P.A.B., S.A.; Formal analysis: A.S., G.E., J.Wan., J.M.R., O.A., S.A.; Funding acquisition: K.G.A., G.G., O.R.-R., A.R., A.V.S.; Investigation: E.D., M.S., N.V.W., T.S.W., D.D., J.Wal., M.S.C., O.K., J.L.M., M.L.; Methodology: A.V.S., A.S., F.A., G.E., J.M.R., J.Wan., E.F.; Resources: F.A.; Software: G.E., S.A., A.S., J.M.R., J.Wan.; Supervision: A.R., A.V.S., F.A., K.G.A., O.R.-R.; Validation: S.A., A.S., G.E.; Visualization: S.A., G.E., A.S., J.Wan.; Writing–original draft: A.S., E.D., S.A., G.E., E.F., F.A., K.G.A., A.V.S., J.M.R., J.Wan., A.R.; Writing–review and editing: G.E., F.A., K.G.A., A.R., A.G., E.F., A.V.S. A.M.T. performed preprocessing and annotation of the early-lung scRNA-seq dataset. **Competing interests:** A.R. is a co-founder and equity holder of Celsius Therapeutics, is an equity holder in Immunitas, and was a scientific advisory board member of ThermoFisher Scientific, Syros Pharmaceuticals, Neogene Therapeutics, and Asimov until 31 July 2020. Since 1 August 2020, A.R. is an employee of Genentech with equity in Roche. G.G. was partially funded by the Paul C. Zamecnik Chair in Oncology at the Massachusetts General Hospital Cancer Center. G.G. receives research funds from IBM and Pharmacyclics and is an inventor on patent applications related to MSMuTect, MSMutSig, MSIDetect, POLYSOLVER, and TensorQTL. G.G. is a founder of and consultant to, and holds privately held equity in, Scorpion Therapeutics. F.A. is an inventor on a patent application related to TensorQTL. F.A. has been an employee of Illumina, Inc., since 8 November 2021. E.D. is an employee of Bristol Myers Squibb. G.E. has been an employee of Genentech since 4 April 2022. O.R.-R. has been an employee of Genentech since 19 October 2020. She has given numerous lectures on the subject of single-cell genomics to a wide variety of audiences and, in some cases, has received remuneration to cover time and costs. O.R.-R. and A.R. are co-inventors on patent applications filed at the Broad Institute of MIT and Harvard related to single-cell genomics. G.E., E.D., O.R.-R., and A.R. are co-applicants on patent WO 2020/232271 relating to the work in this manuscript. **Data and materials availability:** Raw sequence data are available at AnVIL (https://anvil.terra.bio/#workspaces/anvil-datastorage/AnVIL_GTEx_V9_hg38; dbGaP accession phs000424). Gene expression matrices are available from the GTEx Portal (www.gtexportal.org) and the Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1479). The code used in the analysis is available at Zenodo (*179*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4290
Materials and Methods
Supplementary Text
Figs. S1 to S35
Tables S1 to S20
References (*180*–*198*)
MDAR Reproducibility Checklist

Submitted 19 July 2021; accepted 16 March 2022
10.1126/science.abl4290

1848 SOCIETY
AAAS



An estate gift to AAAS

Going all the way back to 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

Today, we are dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy and for the community to speak with one voice to advance science and engineering in the United States and around the world.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science and society – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

1848 SOCIETY
AAAS

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLE SUMMARY

HUMAN CELL ATLAS

Cross-tissue immune cell analysis reveals tissue-specific features in humans

C. Domínguez Conde†, C. Xu†, L. B. Jarvis†, D. B. Rainbow†, S. B. Wells†, T. Gomes, S. K. Howlett, O. Suchanek, K. Polanski, H. W. King, L. Mamanova, N. Huang, P. A. Szabo, L. Richardson, L. Bolt, E. S. Fasouli, K. T. Mahbubani, M. Prete, L. Tuck, N. Richoz, Z. K. Tuong, L. Campos, H. S. Mousa, E. J. Needham, S. Pritchard, T. Li, R. Elmentaite, J. Park, E. Rahmani, D. Chen, D. K. Menon, O. A. Bayraktar, L. K. James, K. B. Meyer, N. Yosef, M. R. Clatworthy, P. A. Sims, D. L. Farber, K. Saeb-Parsy*‡, J. L. Jones*‡, S. A. Teichmann*‡

INTRODUCTION: Immune cells that seed peripheral tissues play a vital role in health and disease, yet most studies of human immunity focus on blood-derived cells. Immune cells adapt to local microenvironments, acquiring distinct features and functional specialization. Dissecting these molecular adaptations through the systematic assessment of cells across the human body promises to transform our understanding of the immune system at the organismal level.

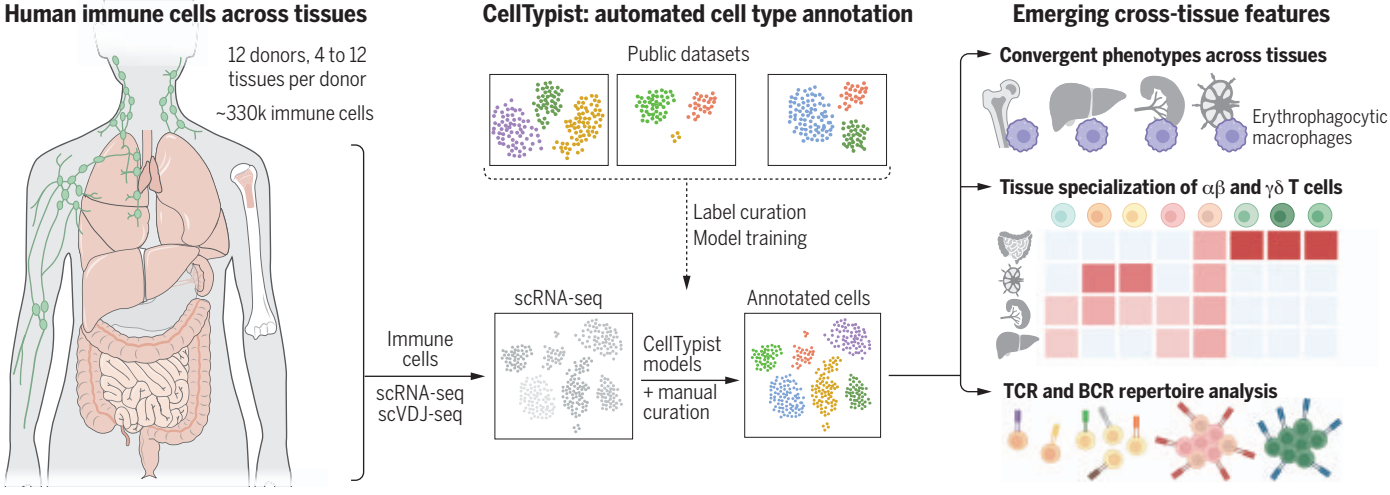
RATIONALE: To comprehensively assess immune cell types, we collected donor-matched tissues from 12 deceased organ donors. We isolated immune cells and performed single-cell RNA sequencing and paired VDJ sequencing for T cell and B cell receptors, resulting in high-quality data for ~330,000 immune cells. To resolve cell identities, we developed CellTypist, a logistic regression–based framework using stochastic gradient descent learning. This cross-tissue annotation enabled interrogation of shared and tissue-specific expression modules and cell states within myeloid and lymphoid cell lineages.

RESULTS: We developed CellTypist by curating and harmonizing public datasets to assemble a comprehensive immune cell type reference database (https://www.celltypist.org). CellTypist was then applied to our data, generated across multiple tissues and individuals (https://www.tissueimmunecellatlas.org/). Altogether, we detected 101 immune populations and performed extensive cross-tissue comparisons for each subset. Although macrophages displayed prominent tissue-restricted features, some convergent features were also detected. For example, macrophages expressing erythrophagocytosis-related genes were widely found in spleen, liver, bone marrow, and lymph nodes. Heterogeneity within defined subpopulations was also observed, such as migratory dendritic cell adaptations. Within adaptive immune lineages, we identified tissue-specific distributions of memory populations. Plasma cells showed a restricted tissue distribution, whereas memory B cells were more widely distributed. Similarly, tissue-resident memory T (T_{RM}) cells were more restricted in distribution compared with central and effector memory T cells. Notably, T_{RM} cells harbored significant diversity, including αβ and γδ lineages, ascertained by VDJ sequencing. Assessment of clonal dynamics revealed the highest clonal expansions in T_{RM} cells and the most frequent clonal sharing between resident and effector memory populations.

CONCLUSION: Here we present an immune cell atlas of myeloid and lymphoid lineages across adult human tissues. We developed CellTypist for automated immune cell annotation and performed an in-depth dissection of cell populations, identifying 101 cell types or states from more than one million cells, including previously underappreciated cell states. We also uncovered convergent phenotypes across tissues within given lineages and described tissue adaptation signatures for a number of cell types, including macrophages and resident memory T cells. Together, we have extended our understanding of how human immunity functions as an integrated, cross-tissue network, and we provide the scientific community with several key new resources. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: st9@sanger.ac.uk (S.A.T.); jls53@medschl.cam.ac.uk (J.L.J.); ks10014@cam.ac.uk (K.S.-P.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as C. Domínguez Conde *et al.*, *Science* **376**, eabl5197 (2022). DOI: 10.1126/science.abl5197

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl5197



Cross-tissue atlas of human immune cells. Donor-matched tissues from 12 deceased organ donors were profiled using single-cell RNA sequencing [including single-cell VDJ sequencing (scVDJ-seq)]. Cell type annotation was achieved with CellTypist, an automated cell type annotation pipeline, followed by manual curation. In-depth analysis of transcriptional features revealed convergent and divergent gene expression programs for each cell lineage across lymphoid and nonlymphoid tissues. scVDJ-seq enabled repertoire analysis across cell subsets and tissues. scRNA-seq, single-cell RNA sequencing; TCR, T cell receptor; BCR, B cell receptor.

RESEARCH ARTICLE

HUMAN CELL ATLAS

Cross-tissue immune cell analysis reveals tissue-specific features in humans

C. Domínguez Conde[1]†, C. Xu[1]†, L. B. Jarvis[2]†, D. B. Rainbow[2]†, S. B. Wells[3]†, T. Gomes[1], S. K. Howlett[2], O. Suchanek[4], K. Polanski[1], H. W. King[5], L. Mamanova[1], N. Huang[1], P. A. Szabo[6], L. Richardson[1], L. Bolt[1], E. S. Fasouli[1], K. T. Mahbubani[7], M. Prete[1], L. Tuck[1], N. Richoz[4], Z. K. Tuong[1,4], L. Campos[1,8], H. S. Mousa[2], E. J. Needham[2], S. Pritchard[1], T. Li[1], R. Elmentaite[1], J. Park[1], E. Rahmani[9,10], D. Chen[3], D. K. Menon[11], O. A. Bayraktar[1], L. K. James[5], K. B. Meyer[1], N. Yosef[9,10,12,13], M. R. Clatworthy[1,4], P. A. Sims[3], D. L. Farber[6], K. Saeb-Parsy[7]*‡, J. L. Jones[2]*‡, S. A. Teichmann[1,14]*‡

Despite their crucial role in health and disease, our knowledge of immune cells within human tissues remains limited. We surveyed the immune compartment of 16 tissues from 12 adult donors by single-cell RNA sequencing and VDJ sequencing generating a dataset of ~360,000 cells. To systematically resolve immune cell heterogeneity across tissues, we developed CellTypist, a machine learning tool for rapid and precise cell type annotation. Using this approach, combined with detailed curation, we determined the tissue distribution of finely phenotyped immune cell types, revealing hitherto unappreciated tissue-specific features and clonal architecture of T and B cells. Our multitissue approach lays the foundation for identifying highly resolved immune cell types by leveraging a common reference dataset, tissue-integrated expression analysis, and antigen receptor sequencing.

The immune system is a dynamic and integrated network made up of many different cell types distributed across the body that act together to maintain tissue homeostasis and mediate protective immunity. In recent years, a growing appreciation of immune ontogeny and diversity across tissues has emerged. For example, we have gained insights into how macrophages derived in embryogenesis contribute to the distinctive adaptation of tissue-resident myeloid cells, such as Langerhans cells in the skin, microglia in the brain, and Kupffer cells in the liver (*1*–*3*). Other populations, such as innate lymphoid cells (ILCs), including natural killer (NK) cells and nonconventional T cells [natural killer T (NKT) cells, mucosal-associated invariant T (MAIT) cells, and gamma-delta (γδ) T cells], have circulating counterparts but are highly enriched at barrier sites where they mediate tissue defense and repair (*4*). In addition, following resolution of an immune response, antigen-specific, long-lived tissue-resident memory T (T_{RM}) cells persist in diverse sites, where they provide optimal protection against secondary infections [reviewed in (*5*–*7*)].

Studies in mice have revealed the central role of tissue immune responses in protective immunity, antitumor immunity, and tissue repair; however, human studies have largely focused on peripheral blood as the most accessible site. Given that circulating immune cells make up only a subset of the entire immune cell landscape, understanding human immunity requires a comprehensive assessment of the properties and features of immune cells within and across tissues. Recent progress in the analysis of tissue immune cells has implemented organ-focused approaches (*8*–*12*), whereas use of tissues obtained from organ donors allows for analysis of immune cells across multiple sites across an individual (*13*–*19*). We previously reported single-cell RNA sequencing (scRNA-seq) analysis of T cells in three tissues from two donors (*20*), identifying tissue-specific signatures. However, despite the effort to assemble murine (*21*) and human (*22*, *23*) multitissue atlases, large-scale cross-tissue scRNA-seq studies that investigate tissue-specific features of innate and adaptive immune compartments have not been reported.

Furthermore, a fundamental challenge of increasingly large single-cell transcriptomics datasets is cell type annotation, including identifying rare cell subsets and distinguishing novel discoveries from previously described cell populations. Currently available automated annotation workflows leverage organ-focused studies and lack a comprehensive catalog of all cell types found across tissues. Therefore, a single unified approach is required to provide an in-depth dissection of immune cell type and immune state heterogeneity across tissues.

To address these needs, we comprehensively profiled immune cell populations isolated from a wide range of donor-matched tissues from 12 deceased individuals, generating nearly 360,000 single-cell transcriptomes. To systematically annotate multitissue immune cells, we developed CellTypist, a machine learning framework for cell type prediction initially trained on data from studies across 20 human tissues (see supplementary text in the supplementary materials) and then updated and extended by integrating immune cells from our dataset.

Results

CellTypist: A pan-tissue immune database and a tool for automated cell type annotation

To systematically assess immune cell type heterogeneity across human tissues, we performed scRNA-seq on 16 different tissues from 12 deceased organ donors (Fig. 1, A and B, and table S1). The tissues studied included primary (bone marrow) and secondary (spleen and lung-draining and mesenteric lymph nodes) lymphoid organs, mucosal tissues (gut and lung), as well as blood and liver. When available, we also included additional donor-matched samples from tissues such as thymus, skeletal muscle, and omentum. Immune cells were isolated from tissues as detailed in the methods. After stringent quality control, we obtained a total of 357,211 high-quality cells, of which 329,762 belonged to the immune compartment (fig. S1, A and B).

Robust cell type annotation remains a major challenge in single-cell transcriptomics. To computationally predict cellular heterogeneity in our dataset, we developed CellTypist (*24*), a cell type database, which in its current form is focused on immune cells, that provides a directly interpretable pipeline for the automatic annotation of scRNA-seq data (Fig. 1C). One of the distinctive and valuable aspects of CellTypist is that its training set encompasses a wide range of immune cell types across tissues. This breadth is of critical importance given that immune compartments are shared across tissues, warranting accurate and automated cell annotation in an organ-agnostic manner. In brief, to develop CellTypist,

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. [2]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. [3]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA. [4]Molecular Immunity Unit, Department of Medicine, University of Cambridge, Cambridge, UK. [5]Centre for Immunobiology, Blizard Institute, Queen Mary University of London, London, UK. [6]Department of Microbiology and Immunology, Columbia University Irving Medical Center, New York, NY, USA. [7]Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, UK. [8]West Suffolk Hospital NHS Trust, Bury Saint Edmunds, UK. [9]Center for Computational Biology, University of California, Berkeley, Berkeley, CA, USA. [10]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, USA. [11]Department of Anaesthesia, University of Cambridge, Cambridge, UK. [12]Chan Zuckerberg Biohub, San Francisco, CA, USA. [13]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [14]Theory of Condensed Matter, Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, UK.
***Corresponding author. Email: st9@sanger.ac.uk (S.A.T.); jls53@medschl.cam.ac.uk (J.L.J.); ks10014@cam.ac.uk (K.S.-P.)**
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

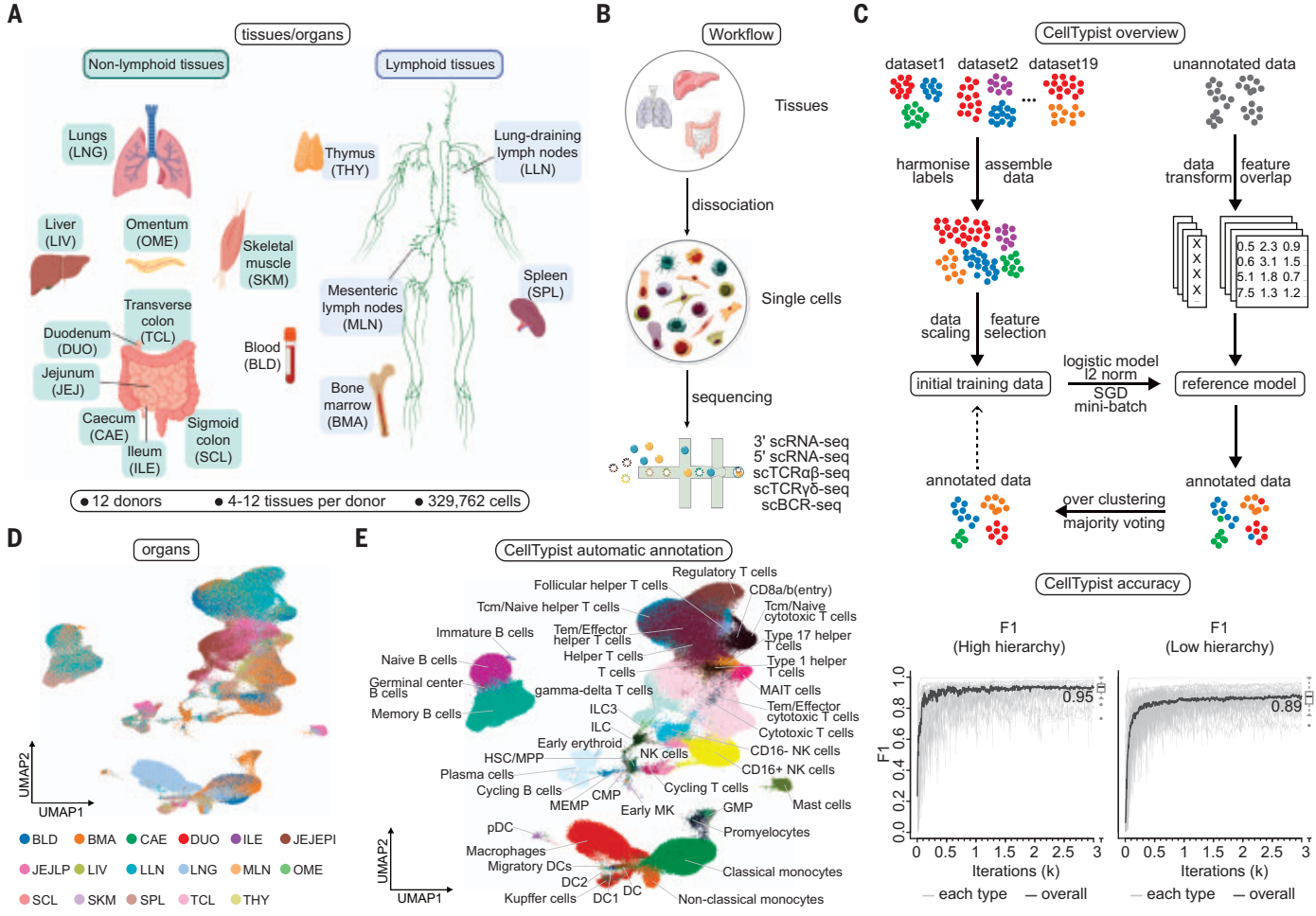


Fig. 1. Automated annotation of immune cells across human tissues using CellTypist. (**A**) Schematic showing sample collections of human lymphoid and nonlymphoid tissues and their assigned tissue name acronyms. (**B**) Schematic of single-cell transcriptome profiling and paired sequencing of αβ TCR, γδ TCR, and BCR variable regions. (**C**) Workflow of CellTypist including data collection, processing, model training, and cell type prediction (upper panel). Performance curves showing the F1 score at each iteration of training with mini-batch stochastic gradient descent for high- and low-hierarchy CellTypist models, respectively (lower panel). The black curve represents the median F1 score averaged across the individual F1 scores of all predicted cell types. (**D**) Uniform manifold approximation and projection (UMAP) visualization of the immune cell compartment colored by tissues. Note that jejunum samples in (A) were further split into epithelial (JEJEPI) and lamina propria (JEJLP) fractions. (**E**) As in (D), but colored by predicted cell types using CellTypist.

we integrated cells from 20 different tissues from 19 reference datasets (fig. S2) where we had deeply curated and harmonized cell types at two knowledge-driven hierarchical levels (figs. S3 to S8). This was followed by a machine learning framework to train a model using logistic regression with stochastic gradient descent learning (see methods). Performance of the derived models, as measured by the precision, recall, and global F1 score, reached ~0.9 for cell type classification at both the high- and low-hierarchy levels (Fig. 1C and fig. S9, A and B). Notably, representation of a given cell type in the training data was a major determinant of its prediction accuracy (fig. S9C), implying that higher model performance can be achieved by incorporating additional datasets. Moreover, CellTypist prediction was robust to differences between training and query datasets including gene expression sparseness (fig. S10) and batch effects (fig. S11).

Next, we applied CellTypist's high-hierarchy (i.e., low-resolution, 32 cell types) classifier to our cross-tissue dataset (Fig. 1D) and found 15 major cell populations (fig. S1C). At this level of resolution, clear compositional patterns emerged in lymphoid versus nonlymphoid tissues, and within the lymphoid tissues between spleen and lymph nodes, and appeared not to be driven by differences in dissociation protocols (fig. S12). As the training datasets of CellTypist contained hematopoietic tissues with definitive annotations for progenitor populations, the classifier was able to resolve several progenitors including hematopoietic stem cells (HSCs), multipotent progenitors (MPPs), promyelocytes, early megakaryocytes, and pre-B and pro-B cells. Furthermore, CellTypist clearly resolved monocytes and macrophages, which often form a transcriptomic continuum in scRNA-seq datasets owing to their functional plasticity. Thus, CellTypist was successfully able to identify major groups of cell populations with different abundances in our dataset (fig. S1C).

To automatically annotate fine-grained immune subpopulations, we next applied the low-hierarchy (high-resolution, 91 cell types and subtypes) classifier, which was able to classify cells into 43 specific subtypes, including subsets of T cells, B cells, ILCs, and mononuclear phagocytes (Fig. 1E). This revealed a high degree of heterogeneity within the T cell compartment, not only distinguishing between αβ and γδ T cells but also between $CD4^+$ and $CD8^+$ T cell subtypes and their more detailed effector and functional phenotypes. Specifically, the $CD4^+$ T cell cluster was classified into helper, regulatory, and cytotoxic subsets,

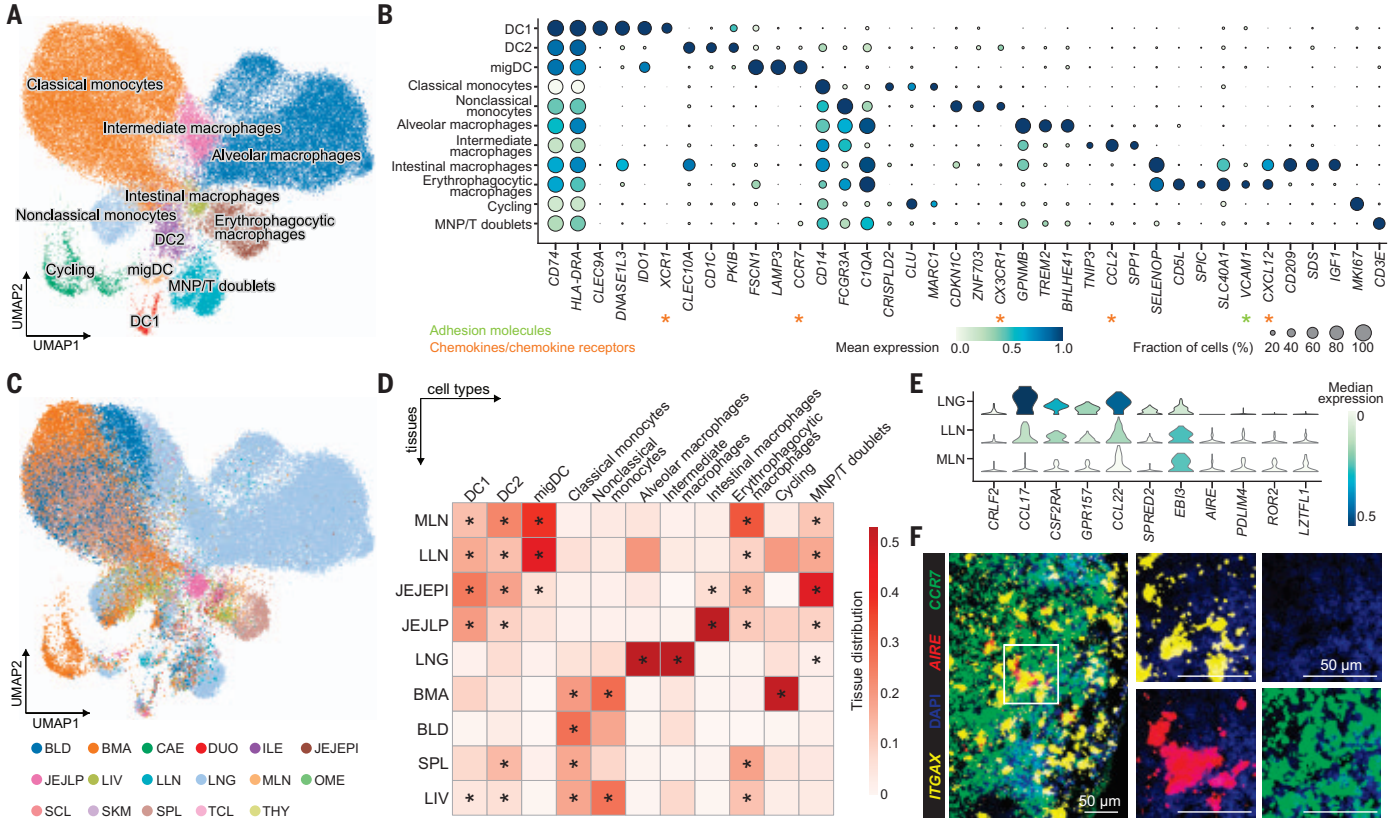


Fig. 2. Myeloid compartment across tissues. (**A**) UMAP visualization of the cell populations in the myeloid compartment. (**B**) Dot plot for expression of marker genes of the identified myeloid populations. Color represents maximum-normalized mean expression of cells expressing marker genes, and size represents the percentage of cells expressing these genes. (**C**) UMAP visualization of the tissue distribution in the myeloid compartment. (**D**) Heatmap showing the distribution of each myeloid cell population across different tissues. Cell numbers are normalized within each tissue and later calculated as proportions across tissues. Only tissues containing >50 myeloid cells in at least two donors were included. Asterisks mark significant enrichment in a given tissue relative to the remaining tissues [Poisson regression stratified by donors, $P < 0.05$ after Benjamini-Hochberg (BH) correction]. (**E**) Violin plot for genes differentially expressed in migratory dendritic cells across tissues. Color represents maximum-normalized mean expression of cells expressing marker genes. (**F**) smFISH visualization of *ITGAX*, *CCR7*, and *AIRE* transcripts, validating the *AIRE*$^+$ migratory dendritic cells in lung-draining lymph nodes.

and the CD8$^+$ T cell clusters contained unconventional T cell subpopulations such as MAIT cells. In the B cell compartment, a clear distinction was observed between naïve and memory B cells. Moreover, CellTypist revealed three distinct subsets of dendritic cells (DCs)—DC1, DC2, and migratory DCs (migDCs) (*25*, *26*)—again highlighting the granularity that CellTypist can achieve. This fine-grained dissection of each compartment also allowed for the cross-validation and consolidation of cell types by examining the expression of marker genes derived from CellTypist models in cells from our dataset (fig. S1D).

To summarize, we generated an in-depth map of immune cell populations across human tissues and developed a framework for automated annotation of immune cell types and subtypes. CellTypist produced fine-grained annotations on our multitissue and multilineage dataset, and its performance, as assessed on multiple metrics, was comparable to or better than other label-transfer methods, with minimal computational cost (figs. S13 and S14). This approach allowed us to identify fine-grained cell subtypes, such as the progenitors and dendritic cell compartments at full transcriptomic breadth, resolving 43 cell states in total across our dataset. This automated annotation forms the basis for further cross-tissue comparisons of cell compartments in the sections below.

Tissue-restricted features of mononuclear phagocytes

Mononuclear phagocytes, including monocytes, macrophages, and dendritic cells, are critical for immune surveillance and tissue homeostasis. Automatic annotation by CellTypist identified eight populations in this compartment (fig. S15A). To explore macrophage heterogeneity further, we built upon CellTypist's annotation by performing additional manual curation, which revealed further heterogeneity within the macrophages (Fig. 2A and fig. S15B). The identities of these cells were supported by expression of well-established marker genes (Fig. 2B) and by markers independently identified from CellTypist models (fig. S15C). Moreover, existence of these cell types was cross-validated, and thus consolidated, in the training datasets of CellTypist (fig. S15D), as well as in myeloid cells from two additional studies of the human intestinal tract (*27*) (fig. S15E) and lung (*28*) (fig. S15F).

Among macrophages, lung-resident cells constituted the majority and were classified into two major clusters: (i) alveolar macrophages expressing *GPNMB* and *TREM2*, markers that have been related to alveolar macrophages (*29*) and disease-associated macrophages (*30*), respectively; and (ii) intermediate macrophages with distinctive expression of *TNIP3* (Fig. 2, B to D). TNIP3 (TNFAIP3-interacting protein 3) binds to A20, encoded by *TNFAIP3*, and inhibits tumor necrosis factor, interleukin-1 (IL-1), and lipopolysaccharide-induced nuclear factor κB (NF-κB) activation (*31*). Its expression in lung macrophages may be related to underlying pathology, as it was primarily detected in one donor (A29), a multitrauma donor with

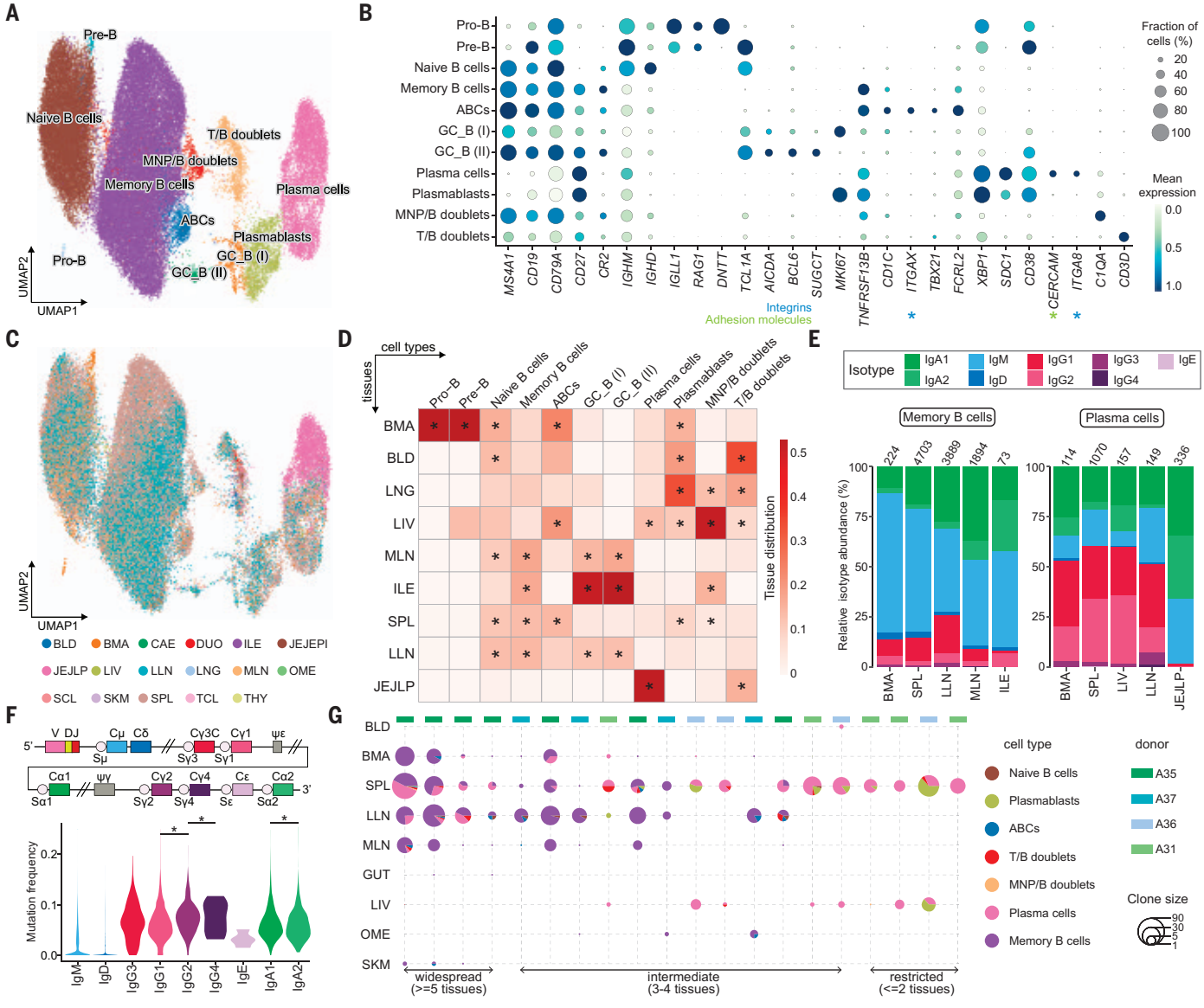


Fig. 3. B cell compartment across tissues. (**A**) UMAP visualization of the cell populations in the B cell compartment. (**B**) Dot plot for expression of marker genes of the identified B cell populations. Color represents maximum-normalized mean expression of cells expressing marker genes, and size represents the percentage of cells expressing these genes. (**C**) UMAP visualization of the tissue distribution in the B cell compartment. (**D**) Heatmap showing the distribution of each B cell population across different tissues. Cell numbers are normalized within each tissue and later calculated as proportions across tissues. Only tissues containing >50 B cells in at least two donors were included. Asterisks mark significant enrichment in a given tissue relative to the remaining tissues (Poisson regression stratified by donors, $P < 0.05$ after BH correction). (**E**) Stacked bar plots showing the isotype distribution per tissue within memory B cells and the plasma cells. (**F**) Violin plot of the hypermutation frequency on the IgH chain across isotypes. Significant difference among IgG4, IgG2, and IgG1 and between IgA2 and IgA1 is marked by asterisks (Wilcoxon rank sum test, $P < 0.05$). (**G**) Scatterpie plot showing the tissue distribution and B cell subsets of expanded clonotypes (>10 cells). Each vertical line represents one clonotype.

substantial lung contusions. Notably, this population also expressed the monocyte-recruiting chemokine CCL2 (Fig. 2B), providing a means of replenishing the lung macrophage pool.

Other macrophage subsets in our data also showed a high degree of tissue restriction (Fig. 2D). Erythrophagocytic macrophages, including red pulp macrophages and Kupffer cells, mainly populated the spleen and liver, as expected, and shared high expression of *CD5L*, *SCL40A1*, and the transcription factor *SPIC* (*32*). Notably, a number of macrophages from lymph nodes clustered together with erythrophagocytic macrophages, pointing to the presence of a small number of iron-recycling macrophages in lymph nodes (Fig. 2D). Another macrophage population specifically present in the gut expressed *CD209* (encoding DC-SIGN) and *IGF1*, markers that have been previously reported in mature intestinal macrophages and M2-like macrophages, respectively (*33*, *34*). Lastly, monocytes were clearly grouped in two major subgroups, classical and nonclassical monocytes, which differed in the expression of *CD14*, *FCGR3A*, and *CX3CR1* and in their tissue distribution (Fig. 2, A to D).

Among dendritic cells, DC1 expressed *XCR1* and *CLEC9A*, consistent with their identity as conventional DCs (DC1), specialized for cross-presentation of antigens (Fig. 2B). Conventional DC2s expressed *CD1C* and *CLEC10A*, and migDCs were *CCR7*$^{+}$*LAMP3*$^{+}$. *CCR7* is up-regulated in

tissue DCs following Toll-like receptor or Fc-gamma receptor ligation (*35*, *36*), enabling migration toward CCL19/21-expressing lymphatic endothelium and stromal cells in the T cell zone of lymph nodes (*37*, *38*). Consistent with this, we observed a marked enrichment of *CCR7*$^{+}$ migDCs in lung-draining and mesenteric lymph nodes (Fig. 2D). Notably, migDCs showed up-regulation of *AIRE*, *PDLIM4*, and *EBI3* in lymph nodes (Fig. 2E). Extrathymic expression of the autoimmune regulator AIRE has been reported in humans (*39*, *40*), however, its functional role in secondary lymphoid organs remains poorly understood and is a matter of intense research (*41*–*43*). We validated the presence of migDCs in lung-draining lymph nodes by immunofluorescence (fig. S16A) and *AIRE* expression by single-molecule fluorescence in situ hybridization (smFISH) (Fig. 2F). In addition, another subpopulation of migDCs found in lung and lung-draining lymph nodes up-regulated *CRLF2* [encoding thymic stromal lymphopoietin receptor (TSLPR)], chemokines (*CCL22* and *CCL17*), *CSF2RA*, and *GPR157* (Fig. 2E). TSLPR is involved in the induction of T helper 2 (T_H2) cell responses in asthma (*44*). Expression of these genes in lung DCs was also observed in published scRNA-seq datasets (*45*, *46*) (fig. S16, B and C). These observations suggest that dendritic cell activation coincides with the acquisition of tissue-specific markers that differ depending on the local microenvironment.

Overall, our analysis of the myeloid compartment has revealed shared and tissue-restricted features of mononuclear phagocytes, including alveolar macrophages in the lung, iron-recycling macrophages mostly localized in the spleen, and subtypes of migratory dendritic cells.

B cell subsets and immunoglobulin repertoires across tissues

B cells comprise progenitors in the bone marrow, developmental states in lymphoid tissues, and terminally differentiated memory and plasma cell states in lymphoid and nonlymphoid tissues. They play a central role in humoral immunity via the production of antibodies tailored to specific body sites. We first annotated the B cells using CellTypist and obtained six populations (fig. S17A). Manual curation revealed further heterogeneity in memory B cells and plasma cells, identifying 11 cell types in total (Fig. 3A and fig. S17B). As with the myeloid compartment, we cross-checked and verified these cell types in CellTypist training datasets (fig. S17, C and D) and in two independent immune datasets from gut (*27*) and lung (*28*) (fig. S17, E and F).

Naïve B cells expressed *TCL1A* and were primarily found in lymphoid tissues (Fig. 3, B to D). In addition, we identified two populations of germinal center B cells, expressing *AICDA* and *BCL6*, that differed in their proliferative states (marked by *MKI67*). Of note, we did not find differential expression of dark zone and light zone marker genes in these two populations, probably reflecting limited germinal center activity in our adult donors. Moreover, these germinal center populations were present in lymph nodes and diverse gut regions (Fig. 3, C and D), presumably representing Peyer's patches. Within memory B cells, which were characterized by expression of the B cell lineage markers (*MS4A1* and *CD19*) and *TNFRSF13B*, we found a transcriptomically distinct cluster positive for *ITGAX*, *TBX21*, and *FCRL2*, encoding CD11c, T-bet, and the Fc receptor-like protein 2, respectively (Fig. 3B). CD11c^{+}T-bet^{+} B cells, also known as age-associated B cells (ABCs), have been reported in autoimmunity and aging (*47*–*49*) and likely correspond to this *ITGAX*$^{+}$ memory B cell population. Unlike conventional memory B cells, ABCs showed relatively low expression of *CR2* (encoding CD21) and *CD27* (Fig. 3B). This subset was mainly present in the liver and bone marrow, while in secondary lymphoid organs, it was primarily found in the spleen, as confirmed by flow cytometry and immunofluorescence (Fig. 3, C and D, and fig. S18). This data deepens our understanding of the phenotype of this nonclassical subtype of memory B cells and their tissue distribution.

We uncovered two populations of plasmablasts and plasma cells marked by expression of *CD38*, *XBP1*, and *SDC1*. Whereas the former expressed *MKI67* and were found in the spleen, liver, bone marrow, and blood, the latter expressed the integrin alpha-8–encoding gene *ITGA8* and the adhesion molecule *CERCAM* and were enriched in the jejunum and liver (Fig. 3, B to D). *ITGA8*$^{+}$ plasma cells have recently been reported in the context of an analysis of bone marrow plasma cells (*50*) and are likely a long-lived plasma cell population that is quiescent and tissue resident. Here we expand their tissue distribution to the liver and gut and describe their specific clonal distribution pattern.

B cells have an additional source of variability due to VDJ recombination, somatic hypermutation, and class switching, which can relate to cell phenotype. Therefore, we performed targeted enrichment and sequencing of B cell receptor (BCR) transcripts to assess isotypes, hypermutation levels, and clonal architecture of the B cell populations identified above. This analysis revealed an isotype and subclass usage pattern that related to cellular phenotype (fig. S19A). As expected, naïve B cells mainly showed a subclass preference for immunoglobulin M (IgM) and IgD. Notably, while evidence of class switching to IgA1 and IgG1 was seen within memory B cells (including ABCs), plasmablasts and plasma cells also showed class switching to IgA2 and IgG2.

To determine to what extent this isotype subclass bias correlated with the tissue of origin, we assessed each cell state independently (requiring a minimum cell count of 50). Memory B cells showed a bias toward IgA1 in the mesenteric lymph nodes, and toward IgA2 in the ileum, where Peyer's patches are found (Fig. 3E). In the plasma cell compartment, we found an even more pronounced preference for IgA2 in the gut region (specifically in the jejunum lamina propria), consistent with the known dominance of this isotype at mucosal surfaces (Fig. 3E). Of note, plasma cells in the bone marrow, liver, and spleen were composed of >20% IgG2 subclass. With more limited numbers, we reported isotype distributions across tissues for other B cell populations (fig. S19, B and C). ABCs were dominated by IgM in both the spleen and lung-draining lymph nodes, consistent with previous reports (*51*).

Somatic hypermutation levels were, as expected, lowest in naïve B cells and highest in plasma cells (fig. S19D). Between isotypes and subclasses, somatic hypermutation did not differ significantly. Nonetheless, there was a tendency toward higher mutation rates in the distal classes IgG2, IgG4, and IgA2, which are downstream from the IgH locus and can thus accumulate more mutations during sequential switching (*52*) (Fig. 3F). We also explored the occurrence of sequential class switching events in our data by assessing the isotype frequency among expanded clonotypes (>10 cells). Mixed isotype clones were rare in our data (fig. S19E).

Next, we evaluated the distribution of expanded clones across tissues and cell types and found three major groups of clones—those present in only two tissues, three to four tissues, or five or more tissues, respectively (Fig. 3G), similar to previously reported patterns of B cell clonal tissue distribution (*53*). While the clones restricted to two tissues, typically between the spleen and the liver or bone marrow, were enriched in plasma cells, those distributed across more than five tissues, including lymph nodes, were over-represented in memory B cells. Together, these findings suggest that tissue-restricted clones may represent long-term immunological memory maintained by long-lived plasma cells resident in the bone marrow and spleen as well as liver in our data.

Overall, we characterized nine cell states in the B cell compartment and gained insights from in-depth characterization of both naïve and memory B cell as well as plasma cell subsets. We identified distinct clonal distribution patterns for the more tissue-restricted long-lived quiescent plasma cells versus the broad tissue distribution of classical memory B cell clones.

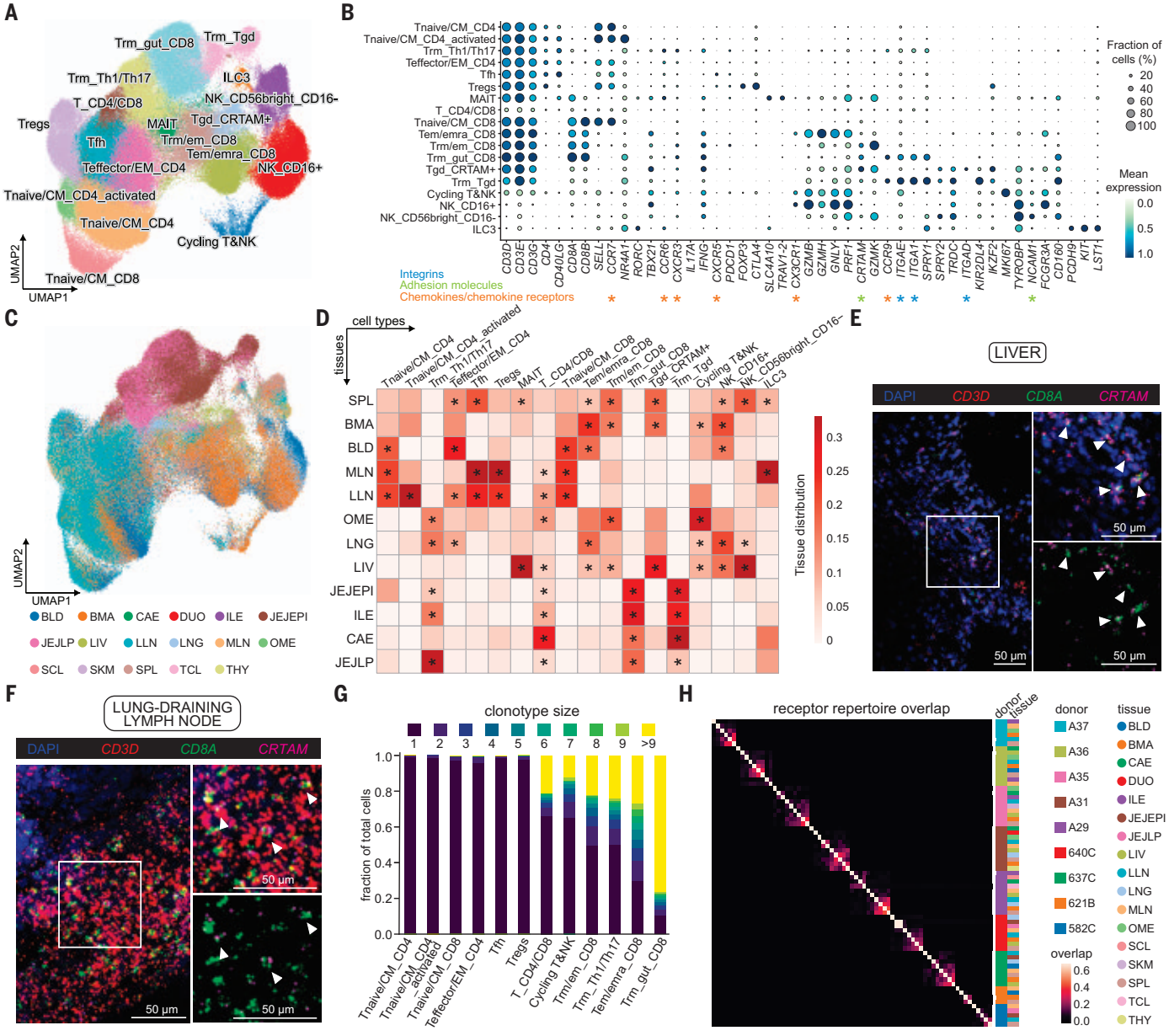


Fig. 4. Tissue compartmentalization and site-specific adaptations of T cells and innate lymphoid cells. (**A**) UMAP visualization of T cells and ILCs across human tissues colored by cell types. (**B**) Dot plot for expression of marker genes of the identified immune populations. Color represents maximum-normalized mean expression of cells expressing marker genes, and size represents the percentage of cells expressing these genes. (**C**) UMAP visualization of T cells and ILCs colored by tissues. (**D**) Heatmap showing the distribution of each T cell or ILC population across different tissues. Cell numbers are normalized within each tissue and later calculated as proportions across tissues. Only tissues containing >50 ILC or T cells in at least two donors were included. Asterisks mark significant enrichment in a given tissue relative to the remaining tissues (Poisson regression stratified by donors, $P < 0.05$ after BH correction). (**E** and **F**) smFISH visualization of *CD3D*, *CD8A*, and *CRTAM* transcripts, validating the tissue-resident memory CD8$^+$ T cell population in the liver and lung-draining lymph nodes. (**G**) TCR repertoire analysis of T cells across tissues. Stacked bar plot shows the fraction of cells in a given cluster binned by clonotype size. (**H**) Heatmap showing the repertoire overlap between expanded clones (>1 cell) across tissues and donors as determined by Jaccard distance.

scRNA-seq analysis of T cells and innate lymphocytes reveals lineage and tissue-specific subsets

For annotation of the T cell and innate lymphocyte compartment, CellTypist identified 18 cell types (fig. S20A). After manual inspection, these clusters were further divided into additional subtypes (e.g., for cytotoxic T cells) (Fig. 4A and fig. S20B). As described above for the myeloid and B cell compartments, identities of the derived cell types were cross-validated in the immune compartment of the CellTypist training datasets (fig. S20, C and D) and the two independent studies of gut (*27*) and lung (*28*) (fig. S20, E and F).

Naïve/central memory CD4$^+$ T cells were transcriptionally close to naïve CD8$^+$ T cells as defined by high expression of *CCR7* and *SELL* and were mainly found in lymphoid sites (Fig. 4B). Other CD4$^+$ T cells identified included T follicular helper (T_{FH}) cells expressing *CXCR5*; regulatory T cells (T_{regs}) expressing *FOXP3* and *CTLA4*; effector memory

$CD4^+$ T cells; and tissue-resident memory T_H1 and T_H17 cells expressing *CCR9*, *ITGAE*, and *ITGA1* found largely in intestinal sites (jejunum and ileum) and lungs (Fig. 4, B to D). Within the memory $CD8^+$ T compartment, we found three major subsets: T_{RM}_gut_CD8 (resident memory T cell, T_{RM}), $T_{EM/EMRA}$_CD8 (effector memory T cell, T_{EM}; effector memory reexpressing CD45RA, T_{EMRA}), and $T_{RM/EM}$_CD8. These subsets were characterized by differential expression of the chemokine receptors *CCR9* and *CX3CR1* and the activation marker *CRTAM* (Fig. 4B). The T_{RM}_gut_CD8 population ($CCR9^+$) expressed the tissue-residency markers *ITGAE* and *ITGA1*, encoding CD103 and CD49a, respectively, and localized to intestinal sites (Fig. 4, B to D). By contrast, the $T_{EM/EMRA}$_CD8 population expressing *CX3CR1* was found in blood-rich sites (blood, bone marrow, lung, and liver) and was excluded from lymph nodes and gut (Fig. 4, C and D), consistent with previous flow cytometry analysis of T_{EMRA} cells (*54*) and results showing $CX3CR1^+CD8^+$ T cells as blood-confined and absent from thoracic duct lymph (*55*). The $T_{RM/EM}$_CD8 population expressed high levels of *CRTAM*, a gene previously shown to be expressed by T_{RM} cells (*56*) and was found in spleen, bone marrow, and to a lesser extent in lymph nodes and lungs. This may be a resident population more prevalent in lymphoid sites (*16*). We validated and mapped the $T_{RM/EM}$_CD8 population using smFISH in the liver (Fig. 4E) and lung-draining lymph nodes (Fig. 4F). Furthermore, we validated all three memory $CD8^+$ T cell populations at the protein level using flow cytometry of cells purified from human spleen (fig. S21). Although we could validate *CRTAM* at the RNA level by smFISH, the protein could not be detected without stimulation, suggesting that CRTAM is subject to posttranslational regulation upon T cell receptor (TCR) activation. These three distinct populations represent different states of tissue adaptation and maturation between effector memory and tissue-resident T cell memory states.

We also detected invariant T cell subsets such as MAIT cells, characterized by expression of *TRAV1-2* and *SLC4A10*, and two populations of γδ T cells: T_{RM}_T_{GD} and T_{GD}_$CRTAM^+$. The $CCR9^+$ T_{RM}_T_{GD} population populated the gut and expressed the tissue-residency markers *ITGAE* and *ITGA1*, whereas the T_{GD}_$CRTAM^+$ population overexpressed *CRTAM*, *IKZF2* (encoding HELIOS), and the integrin molecule *ITGAD* (encoding CD11d) and was found primarily in the spleen, bone marrow, and liver (Fig. 4, B to D, and fig. S22, A and B). We validated the latter population by quantitative PCR (qPCR) of flow-sorted $CD3^+TCR\gamma\delta^+$ and $CD3^+TCR\alpha\beta^+$ cells from cryopreserved spleen samples from three donors (fig. S22, C and D). As a small fraction of αβ T cells, marked by low expression of CD52 and CD127, were also noted to express *ITGAD*, the $CD3^+TCR\alpha\beta$ population was split into $CD52^-CD127^-$ and $CD52^+CD127^+$ subpopulations. In keeping with our scRNA-seq data, *ITGAD* expression was high in $CD3^+TCR\gamma\delta$ and $CD52^-CD127^-CD3^+$ $TCR\alpha\beta$, providing additional evidence for the specific expression of this integrin alpha subunit in this subpopulation of γδ T cells.

Lastly, NK cells in our data were represented by two clusters with high expression of either *FCGR3A* (encoding CD16) or *NCAM1* (encoding CD56). We also defined an ILC3 population within a small cluster mixed with NK cells, via expression of markers including *PCDH9* (Fig. 4, A and B). Analyses of the tissue distribution of these populations revealed that, whereas most of the $CD4^+$ T and ILC3 cells were located in the lymph nodes and to some extent in the spleen, cytotoxic T and NK cells were more abundant in the bone marrow, spleen, and nonlymphoid tissues (Fig. 4, C and D).

TCR repertoire analysis shows clonal expansion and distribution patterns within and across tissues

To understand T cell–mediated protection in more depth, we analyzed T cell clonal distribution in a subset of the data within different tissues of a single individual and across different individuals. Chain pairing analysis showed that cells from the T cell clusters mostly contained a single pair of TCRαβ chains (50 to 60%), with orphan (5 to 20%) and extra (5 to 10%) chains present in small fractions of cells (fig. S23A). Notably, the frequency of extra α chains (extra VJ) was higher than that of β chains (extra VDJ), potentially because of more stringent and multilayered allelic exclusion mechanisms at the TCRβ locus compared with TCRα (*57*). As expected, the NK and ILC clusters held no productive TCR chains. Within the γδ T cell clusters, only a small proportion had a productive TCR chain, which may result from cytotoxic T cells co-clustering with γδ T cells. We also carried out γδ TCR sequencing in selected spleen, bone marrow, and liver samples. The γδ TCR sequencing data was subjected to a customized analysis pipeline that we developed and optimized using cellranger followed by contig reannotation with dandelion (see materials and methods), facilitating the full recovery of γδ chains in our data. This analysis confirmed that most of the productive γδ TCR chains originated from the *ITGAD*-expressing γδ T cells (fig. S23B), supporting the robust identification of this population. The T_{RM}_T_{GD} population could not be confirmed by γδ TCR sequencing owing to the lack of sample availability.

We next examined V(D)J gene usage in relation to T cell identity. In the MAIT population, we detected significant enrichment of *TRAV1-2*, as expected (fig. S23C). Selecting only the $TRAV1\text{-}2^+$ cells (MAIT cluster and other clusters) revealed a notable tissue-specific distribution of *TRAJ* segments, with *TRAJ33* in spleen and liver, *TRAJ12* in liver, and *TRAJ29/TRAJ36* in jejunum (fig. S23D). This suggests that there may be different antigens for MAIT cells in the spleen, liver, and gut corresponding to the different metabolomes in these tissues. In addition, full analysis of the TCR repertoire of the MAIT cells revealed previously unappreciated diversity of V segment usage in the β chain (fig. S23D).

We then defined clonally related cells on the basis of identical CDR3 nucleotide sequences to investigate their TCR repertoires. Using this approach, we found that clonally expanded cells were primarily from the resident memory T cell compartment, including the T_H1 and T_H17 populations mentioned above (Fig. 4G and fig. S23E). As expected, these clonotypes were restricted to single individuals, and within an individual they were distributed across tissues and subsets (Fig. 4H and fig. S23, F to H). We found a small number of isolated $CD4^+$ T cell clones that shared T_{regs} and effector T cell phenotypes, possibly owing to low levels of plasticity or to (trans)differentiation from the same naïve precursor cell in the periphery (fig. S23H). Focusing on the most expanded clonotypes (>20 cells), many were widespread across five or more tissues, supporting the systemic nature of tissue-resident immune memory (Fig. 4G). Moreover, we found that several clonotypes present across tissues consisted of a mixture of cells from the $T_{EM/EMRA}$_CD8 and $T_{RM/EM}$_CD8 populations (fig. S23H), suggesting that a single naïve $CD8^+$ T cell precursor can give rise to diverse cytotoxic T cell states, which harbor immune memory across multiple nonlymphoid tissues, emphasizing the plasticity of phenotype and location within a clone.

In summary, we have described 18 T or innate cell states in our data by integrating CellTypist logistic regression models, manual curation, and V(D)J sequencing. This analysis has yielded insights into the MAIT cell compartment and its antigen receptor repertoire distribution that differed between spleen, liver, and gut. For the cytotoxic T cell memory compartment, there was broad sharing of clones across gut regions for T_{RM}_gut_CD8, and mixed $T_{EM/EMRA}$_CD8 and $T_{RM/EM}$_CD8 T cell clonotypes with broad tissue distributions.

A cross-tissue updatable reference of immune cell types and states

After focusing on individual immune compartments, we next took a combined approach to better understand the immune landscape of selected tissues. As shown in Fig. 5A, each tissue has its own immune neighborhood, for

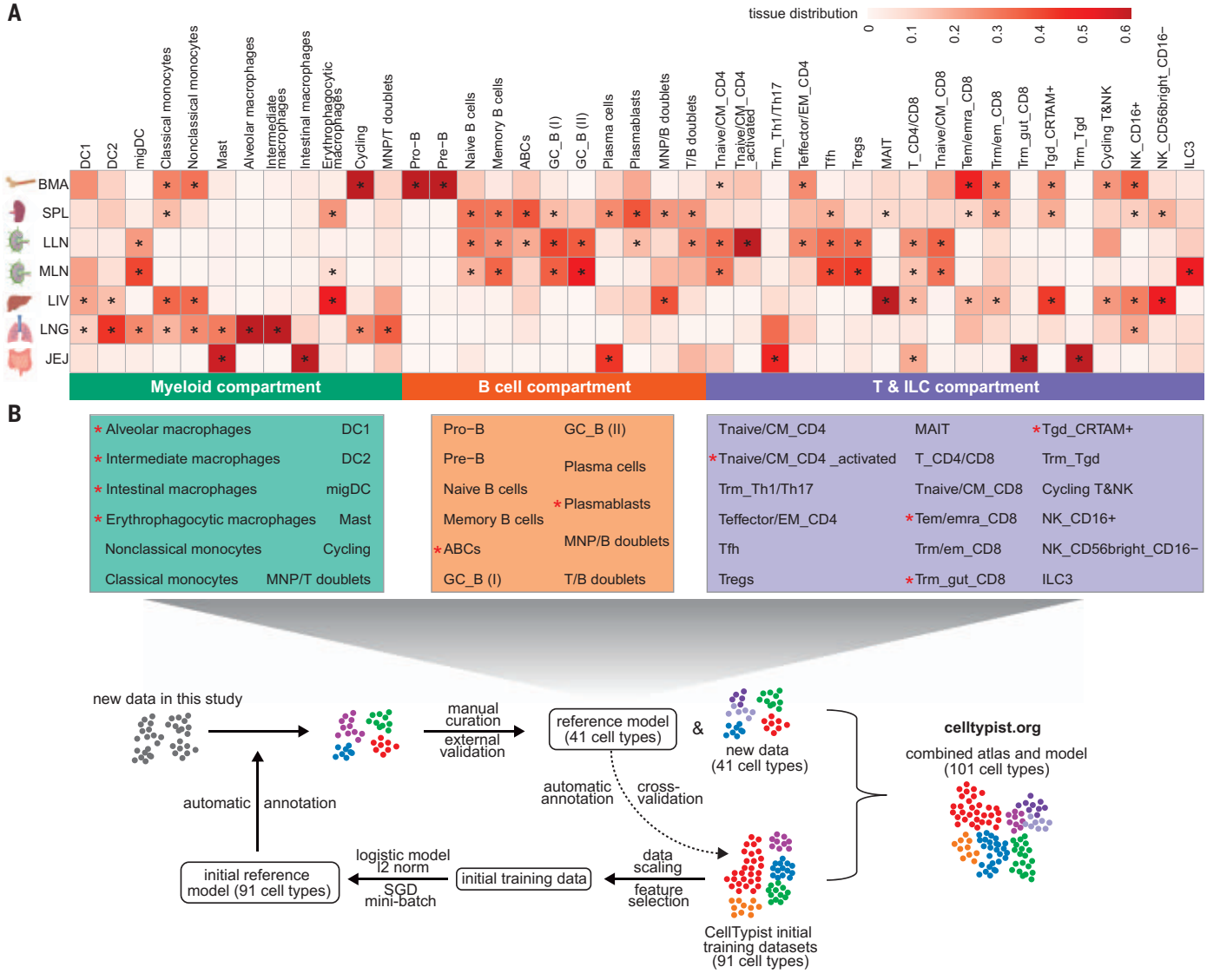


Fig. 5. A cross-tissue updatable reference of immune cell types and cell states. (**A**) Heatmap showing the distribution of manually curated cell types across selected tissues. Cell numbers are normalized within each tissue and later calculated as proportions across tissues. Asterisks mark significant enrichment in a given tissue relative to the remaining tissues (Poisson regression stratified by donors, $P < 0.05$ after BH correction). (**B**) Workflow for the iterative update of CellTypist through the periodic incorporation of curated cell type labels.

example, while spleen and lymph nodes are rich in B cells, composition of their myeloid compartment varies. In particular, a large population of erythrophagocytic macrophages, known as red pulp macrophages, are evident in the spleen (in keeping with their role in red blood cell turnover), whereas lymph nodes are rich in dendritic cells. As expected, bone marrow uniquely contains progenitor populations. Furthermore, lung and liver contain notable numbers of monocytes, including *CX3CR1*$^{+}$ nonclassical monocytes, whereas these cells are absent from the jejunum, perhaps reflecting different degrees of vascularization. In contrast, the jejunum has an abundance of resident memory T cells (CD8^{+} T cells, T_H1 cells, and T_H17 cells) as well as plasma cells.

Our long-term vision for CellTypist is to provide a reference atlas with deeply curated cell types publicly available to the community. Therefore, via a semiautomatic process, we fed the identities of the 41 immune cell types identified in our dataset (including both shared and novel cell type labels) back into CellTypist, demonstrating how CellTypist can be updated and improved over time. Combined with the initial 91 cell types and states included in the reference datasets, CellTypist now comprises 101 annotated cell types (Fig. 5B).

Discussion

Here, we present a multitissue study of immune cells across the human body in diverse organ donors. By sampling multiple organs from the same individuals, which allows for robust control of age, sex, medical history, drug exposure and sampling backgrounds, we reveal tissue-specific expression patterns across the myeloid and lymphoid compartments.

We also introduce CellTypist, a publicly available and updatable framework for automated immune cell type annotation that, in addition to identifying major cell types, can perform fine-grained cell subtype annotation—normally a time-consuming process that requires expert knowledge. We developed CellTypist by integrating and curating data obtained from 19 studies performed across a range of tissues, with in-depth immune cell analysis of 91 harmonized cell type labels. However, as demonstrated here, for example

in the γδ T cell compartment, manual curation after automated annotation still has a role to play in revealing specific cell subtypes that may be absent from the training set. To reduce the need for this manual curation, in the longer term, the CellTypist models will be periodically updated and extended to include further immune and nonimmune subpopulations as more data become available.

Within the myeloid compartment, macrophages showed the most prominent features of tissue specificity. Erythrophagocytic macrophages in the liver and spleen shared features related to iron recycling (*58*) with macrophages in other locations, such as the mesenteric lymph nodes, suggesting that macrophages participate in iron metabolism across a range of tissues. In addition, we characterized subsets of migratory dendritic cells (*CCR7*$^+$) revealing specific expression of *CRLF2*, *CSF2RA*, and *GPR157* in the lung and lung-draining lymph nodes and expression of *AIRE* in the mesenteric and lung-draining lymph nodes. These migratory dendritic cell states are interesting targets for future in-depth functional characterization in the context of allergy, asthma, and other related pathologies (*59*, *60*).

In the lymphoid compartment, we combined single-cell transcriptome and VDJ analysis, which allowed the phenotype of adaptive immune cells to be dissected using complementary layers of single-cell genomics data. Of note, we detected a subset of memory B cells expressing *ITGAX* (CD11c) and *TBX21* (T-bet) that resemble ABCs previously reported to be expanded in aging (*48*), following malaria vaccination (*61*) and in systemic lupus erythematosus patients (*62*). In our data, these B cells did not show clonal expansion, and at least 50% showed IgM subclass, suggesting that they may be present at low levels in healthy tissues and expand upon challenge as well as aging. BCR analysis revealed isotype usage biased toward IgA2 in gut plasma cells, which may be related to structural differences (*63*) or higher resistance to microbial degradation as compared with IgA1 (*64*).

In the T cell compartment, our results provided insights into the heterogeneity of T cell subtypes and their tissue adaptations. Notably, we identified subsets of CD4$^+$ T_{RM} on the basis of functional capacity for interferon-γ or IL-17 production that were mostly localized to intestinal sites, analogous to mouse CD4$^+$ T_{RM} generated from IL-17–producing effector T cells in the gut (*65*). We also identified different subsets of CD8$^+$ T_{RM} including a gut-adapted subset expressing *CCR9*, which mediates homing to intestinal sites via binding to CCL25 (*66*) and another T_{RM}-like subset more targeted to lymphoid sites. TCR clone sharing between memory subtypes of CD8$^+$ T cells suggests their origin from a common precursor, or their differentiation or conversion during migration or maintenance, such as conversion of T_{EM} cells to T_{RM} cells (*56*). We also identified distinct subsets of γδ T cells on the basis of tissue-specific gene expression patterns, showing distinct integrin gene expression and tissue distributions.

In summary, using this dataset of nearly 360,000 single-cell transcriptomes (of which ~330,000 were immune cells) from donor-matched tissues from 12 deceased individuals, we have shown how a combination of CellTypist-based automated annotation, expert-driven cluster analysis, and antigen receptor sequencing can synergize to dissect specific and functionally relevant aspects of immune cells across the human body. We have revealed previously unrecognized features of tissue-specific immunity in the myeloid and lymphoid compartments, and have provided a comprehensive framework for future cross-tissue cell type analysis. Further investigation of human tissue-resident immunity is needed to determine the effect of important covariates such as underlying critical illness, donor age, and gender, as well as considering the immune cell activation status, to gain a defining picture of how human biology influences immune functions. Our deeply characterized cross-tissue immune cell dataset has implications for the engineering of cells for therapeutic purposes and addressing cells to intended tissue locations, as well as for understanding tissue-specific features of infection as well as distinct modes of vaccine delivery to tissues.

Materials and methods summary

Tissue acquisition, processing, and single-cell sequencing

Tissue was obtained from deceased organ donors through the Cambridge Biorepository for Translational Medicine (CBTM, https://www.cbtm.group.cam.ac.uk/), REC 15/EE/0152. Detailed sample locations taken can be found in Fig. 1, and protocols are described in detail in supplementary materials. Additional tissue samples were from Columbia University and were obtained from deceased organ donors at the time of organ acquisition for clinical transplantation through an approved protocol and material transfer agreement with LiveOnNY.

Six donors were processed with a uniform protocol at Cambridge University, where solid tissues were cut into small pieces and then homogenized with two rounds of enzymatic digestion at 37°C for 15 min with mixing. The remaining six donors were subjected to a tissue-adapted protocol with the aim of improving immune cell recovery, and this protocol was harmonized as closely as possible between the two collection sites.

For scRNA-seq experiments, single cells were loaded onto the channels of a Chromium chip (10x Genomics). cDNA synthesis, amplification, and sequencing libraries were generated using either the Single Cell 5′ Reagent (v1 and v2) (Cambridge University) or 3′ Reagent (v3) (Columbia University) Kit. TCRαβ, BCR, and TCRγδ paired VDJ libraries were prepared from samples made with the 5′ Reagent kit. All libraries were sequenced on either a HiSeq 4000 or NovaSeq 6000 instrument.

scRNA-seq and scVDJ-seq data analysis

scRNA-seq data was aligned and quantified using the cellranger software (version 6.1.1, 10x Genomics). Cells from hashtagged samples were demultiplexed using Hashsolo (*67*). Cells with fewer than 1000 unique molecular identifier counts and 600 detected genes were excluded. Doublets were detected using Scrublet (*68*). Downstream analysis from data normalization to graph-based clustering were performed using Scanpy (version 1.6.0) (*69*), with details described in supplementary materials. Data integration was done using BBKNN (*70*) and scVI (*71*), and the results were compared using kBET (*72*).

Single-cell TCR sequencing and single-cell BCR sequencing data were aligned and quantified using the cellranger-vdj software (versions 2.1.1 and 4.0, respectively). For TCRγδ, we implemented a customized pipeline (https://sc-dandelion.readthedocs.io/en/latest/notebooks/gamma_delta.html) owing to cellranger being tuned toward alpha/beta TCR chains. Single-cell TCR sequencing analysis including productive TCR chain pairing, and clonotype detection was performed using the scirpy package (*73*).

CellTypist

Details of CellTypist, including cross-data cell type label harmonization and automated cell annotation, can be found in the supplementary text in the supplementary materials. Briefly, immune cells from 20 tissues of 19 studies were collected and harmonized into consistent labels. These cells were split into equal-sized mini-batches, and these batches were sequentially trained by the l2-regularized logistic regression using stochastic gradient descent learning. Feature selection was performed to choose the top 300 genes from each cell type, and the union of these genes was supplied as the input for a second round of training.

Single-molecule FISH, flow cytometry, qPCR, and immunofluorescence

For smFISH, samples were run using the RNAscope 2.5 LS fluorescent multiplex assay (automated). Slides were imaged on the Perkin Elmer Opera Phenix High-Content Screening System, in confocal mode with 1-μm z-step size, using 20× (NA 0.16, 0.299 μm per pixel) and

40× (NA 1.1, 0.149 μm per pixel) water-immersion objectives.

For flow cytometry, mononuclear cells were either stained ex vivo or after activation with phorbol 12-myristate 13-acetate/ionomycin for 2 hours. Cells were stained with the live/dead marker Zombie Aqua for 10 min at room temperature and then washed with phosphate-buffered saline (PBS) and 0.5% fetal calf serum (FCS), with the CD8 and B cell panels of antibodies.

qPCR was performed in three spleen samples. Cells were stained with the live/dead marker Zombie Aqua for 10 min at room temperature and then washed with PBS +0.5%FCS, followed by staining with the antibodies at 4°C for 45 min. Cell sorting was performed on a BD Fusion 4 laser sorter, and RNA was extracted using a Zymo Research RNA micro kit.

For immunofluorescence, samples were fixed in 1% paraformaldehyde for 24 hours followed by 8 hours in 30% sucrose in PBS and were stained for 2 hours at room temperature with the appropriate antibodies, washed three times in PBS, and mounted in Fluoromount-G (Southern Biotech). Images were acquired using a TCS SP8 (Leica, Milton Keynes, UK) confocal microscope.

REFERENCES AND NOTES

1. D.-M. Popescu *et al*., Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019). doi: 10.1038/s41586-019-1652-y; pmid: 31597962
2. L. C. Davies, S. J. Jenkins, J. E. Allen, P. R. Taylor, Tissue-resident macrophages. *Nat. Immunol.* **14**, 986–995 (2013). doi: 10.1038/ni.2705; pmid: 24048120
3. E. Gomez Perdiguero *et al*., Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature* **518**, 547–551 (2015). doi: 10.1038/nature13989; pmid: 25470051
4. X. Fan, A. Y. Rudensky, Hallmarks of tissue-resident lymphocytes. *Cell* **164**, 1198–1211 (2016). doi: 10.1016/j.cell.2016.02.048; pmid: 26967286
5. D. J. Topham, E. C. Reilly, Tissue-resident memory CD8$^+$ T cells: From phenotype to function. *Front. Immunol.* **9**, 515 (2018). doi: 10.3389/fimmu.2018.00515; pmid: 29632527
6. D. Masopust, A. G. Soerens, Tissue-resident T cells and other resident leukocytes. *Annu. Rev. Immunol.* **37**, 521–546 (2019). doi: 10.1146/annurev-immunol-042617-053214; pmid: 30726153
7. P. A. Szabo, M. Miron, D. L. Farber, Location, location, location: Tissue resident memory T cells in mice and humans. *Sci. Immunol.* **4**, eaas9673 (2019). doi: 10.1126/sciimmunol.aas9673; pmid: 30952804
8. N. Aizarani *et al*., A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* **572**, 199–204 (2019). doi: 10.1038/s41586-019-1373-2; pmid: 31292543
9. J. Zhao *et al*., Single-cell RNA sequencing reveals the heterogeneity of liver-resident immune cells in human. *Cell Discov.* **6**, 22 (2020). doi: 10.1038/s41421-020-0157-z; pmid: 32351704
10. J.-E. Park *et al*., A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, eaay3224 (2020). doi: 10.1126/science.aay3224; pmid: 32079746
11. K. R. James *et al*., Distinct microbial and immune niches of the human colon. *Nat. Immunol.* **21**, 343–353 (2020). doi: 10.1038/s41590-020-0602-z; pmid: 32066951
12. B. J. Stewart *et al*., Spatiotemporal immune zonation of the human kidney. *Science* **365**, 1461–1466 (2019). doi: 10.1126/science.aat5031; pmid: 31604275
13. M. T. Wong *et al*., A high-dimensional atlas of human T cell diversity reveals tissue-specific trafficking and cytokine signatures. *Immunity* **45**, 442–456 (2016). doi: 10.1016/j.immuni.2016.07.007; pmid: 27521270
14. P. Dogra *et al*., Tissue determinants of human NK cell development, function, and residence. *Cell* **180**, 749–763.e13 (2020). doi: 10.1016/j.cell.2020.01.022; pmid: 32059780
15. N. A. Yudanin *et al*., Spatial and temporal mapping of human innate lymphoid cells reveals elements of tissue specificity. *Immunity* **50**, 505–519.e4 (2019). doi: 10.1016/j.immuni.2019.01.012; pmid: 30770247
16. B. V. Kumar, T. J. Connors, D. L. Farber, Human T cell development, localization, and function throughout life. *Immunity* **48**, 202–213 (2018). doi: 10.1016/j.immuni.2018.01.007; pmid: 29466753
17. D. L. Farber, Tissues, not blood, are where immune cells function. *Nature* **593**, 506–509 (2021). doi: 10.1038/d41586-021-01396-y; pmid: 34035530
18. S. P. Weisberg, B. B. Ural, D. L. Farber, Tissue-specific immunity for a changing world. *Cell* **184**, 1517–1529 (2021). doi: 10.1016/j.cell.2021.01.042; pmid: 33740452
19. M. M. L. Poon, D. L. Farber, The whole body as the system in systems immunology. *iScience* **23**, 101509 (2020). doi: 10.1016/j.isci.2020.101509; pmid: 32920485
20. P. A. Szabo *et al*., Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019). doi: 10.1038/s41467-019-12464-3; pmid: 31624246
21. Tabula Muris Consortium, Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* **562**, 367–372 (2018). doi: 10.1038/s41586-018-0590-4; pmid: 30283141
22. S. He *et al*., Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biol.* **21**, 294 (2020). doi: 10.1186/s13059-020-02210-0; pmid: 33287869
23. X. Han *et al*., Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020). doi: 10.1038/s41586-020-2157-4; pmid: 32214235
24. C. Xu, Prete, Teichlab/celltypist: celltypist 0.1.6, Zenodo (2021); https://zenodo.org/record/5736560.
25. C. C. Brown *et al*., Transcriptional basis of mouse and human dendritic cell heterogeneity. *Cell* **179**, 846–863.e24 (2019). doi: 10.1016/j.cell.2019.09.035; pmid: 31668803
26. A.-C. Villani *et al*., Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017). doi: 10.1126/science.aah4573; pmid: 28428369
27. R. Elmentaite *et al*., Cells of the human intestinal tract mapped across space and time. *Nature* **597**, 250–255 (2021). doi: 10.1038/s41586-021-03852-1; pmid: 34497389
28. E. Madissoon *et al*., A spatial multi-omics atlas of the human lung reveals a novel immune cell survival niche. bioRxiv 2021.11.26.470108 [Preprint] (2021). https://doi.org/10.1101/2021.11.26.470108.
29. E. Evren *et al*., Distinct developmental pathways from blood monocytes generate human lung macrophage diversity. *Immunity* **54**, 259–275.e7 (2021). doi: 10.1016/j.immuni.2020.12.003; pmid: 33382972
30. A. Deczkowska, A. Weiner, I. Amit, The physiology, pathology, and potential therapeutic applications of the TREM2 signaling pathway. *Cell* **181**, 1207–1217 (2020). doi: 10.1016/j.cell.2020.05.003; pmid: 32531244
31. A. Martens, G. van Loo, A20 at the crossroads of cell death, inflammation, and autoimmunity. *Cold Spring Harb. Perspect. Biol.* **12**, a036418 (2020). doi: 10.1101/cshperspect.a036418; pmid: 31427375
32. M. Haldar *et al*., Heme-mediated SPI-C induction promotes monocyte differentiation into iron-recycling macrophages. *Cell* **156**, 1223–1234 (2014). doi: 10.1016/j.cell.2014.01.069; pmid: 24630724
33. A. Bujko *et al*., Transcriptional and functional profiling defines human small intestinal macrophage subsets. *J. Exp. Med.* **215**, 441–458 (2018). doi: 10.1084/jem.20170057; pmid: 29273642
34. O. Spadaro *et al*., IGF1 shapes macrophage activation in response to immunometabolic challenge. *Cell Rep.* **19**, 225–234 (2017). doi: 10.1016/j.celrep.2017.03.046; pmid: 28402847
35. F. Sallusto *et al*., Rapid and coordinated switch in chemokine receptor expression during dendritic cell maturation. *Eur. J. Immunol.* **28**, 2760–2769 (1998). doi: 10.1002/(SICI)1521-4141(199809)28:09<2760::AID-IMMU2760>3.0.CO;2-N; pmid: 9754563
36. M. R. Clatworthy *et al*., Immune complexes stimulate CCR7-dependent dendritic cell migration to lymph nodes. *Nat. Med.* **20**, 1458–1463 (2014). doi: 10.1038/nm.3709; pmid: 25384086
37. G. J. Randolph, V. Angeli, M. A. Swartz, Dendritic-cell trafficking to lymph nodes through lymphatic vessels. *Nat. Rev. Immunol.* **5**, 617–628 (2005). doi: 10.1038/nri1670; pmid: 16056255
38. J. G. Cyster, Chemokines and the homing of dendritic cells to the T cell areas of lymphoid organs. *J. Exp. Med.* **189**, 447–450 (1999). doi: 10.1084/jem.189.3.447; pmid: 9927506
39. J. R. Fergusson *et al*., Maturing human CD127+ CCR7+ PDL1+ dendritic cells express AIRE in the absence of tissue restricted antigens. *Front. Immunol.* **9**, 2902 (2019). doi: 10.3389/fimmu.2018.02902; pmid: 30692988
40. P. L. Poliani *et al*., Human peripheral lymphoid tissues contain autoimmune regulator-expressing dendritic cells. *Am. J. Pathol.* **176**, 1104–1112 (2010). doi: 10.2353/ajpath.2010.090956; pmid: 20093495
41. J. Wang *et al*., Single-cell multiomics defines tolerogenic extrathymic Aire-expressing populations with unique homology to thymic epithelium. *Sci. Immunol.* **6**, eabl5053 (2021). doi: 10.1126/sciimmunol.abl5053; pmid: 34767455
42. J. M. Gardner *et al*., Deletional tolerance mediated by extrathymic Aire-expressing cells. *Science* **321**, 843–847 (2008). doi: 10.1126/science.1159407; pmid: 18687966
43. J. M. Gardner *et al*., Extrathymic *Aire*-expressing cells are a distinct bone marrow-derived population that induce functional inactivation of CD4$^+$ T cells. *Immunity* **39**, 560–572 (2013). doi: 10.1016/j.immuni.2013.08.005; pmid: 23993652
44. B. N. Lambrecht, H. Hammad, Biology of lung dendritic cells at the origin of asthma. *Immunity* **31**, 412–424 (2009). doi: 10.1016/j.immuni.2009.08.008; pmid: 19766084
45. F. A. Vieira Braga *et al*., A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* **25**, 1153–1163 (2019). doi: 10.1038/s41591-019-0468-5; pmid: 31209336
46. W. Sungnak *et al*., SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020). doi: 10.1038/s41591-020-0868-6; pmid: 32327758
47. M.-L. Golinski *et al*., CD11c$^+$ B cells are mainly memory cells, precursors of antibody secreting cells in healthy donors. *Front. Immunol.* **11**, 32 (2020). doi: 10.3389/fimmu.2020.00032; pmid: 32158442
48. E. Blanco *et al*., Age-associated distribution of normal B-cell and plasma cell subsets in peripheral blood. *J. Allergy Clin. Immunol.* **141**, 2208–2219.e16 (2018). doi: 10.1016/j.jaci.2018.02.017; pmid: 29505809
49. C. Wei *et al*., A new population of cells lacking expression of CD27 represents a notable component of the B cell memory compartment in systemic lupus erythematosus. *J. Immunol.* **178**, 6624–6633 (2007). doi: 10.4049/jimmunol.178.10.6624; pmid: 17475894
50. A. Kassambara *et al*., GenomicScape: An easy-to-use web tool for gene expression data analysis. Application to investigate the molecular events in the differentiation of B cells into plasma cells. *PLOS Comput. Biol.* **11**, e1004077 (2015). doi: 10.1371/journal.pcbi.1004077; pmid: 25633866
51. M. P. Cancro, Age-associated B cells. *Annu. Rev. Immunol.* **38**, 315–340 (2020). doi: 10.1146/annurev-immunol-092419-031130; pmid: 31986068
52. K. Kitaura *et al*., Different somatic hypermutation levels among antibody subclasses disclosed by a new next-generation sequencing-based antibody repertoire analysis. *Front. Immunol.* **8**, 389 (2017). doi: 10.3389/fimmu.2017.00389; pmid: 28515723
53. W. Meng *et al*., An atlas of B-cell clonal distribution in the human body. *Nat. Biotechnol.* **35**, 879–884 (2017). doi: 10.1038/nbt.3942; pmid: 28829438
54. J. J. C. Thome *et al*., Spatial map of human T cell compartmentalization and maintenance over decades of life. *Cell* **159**, 814–828 (2014). doi: 10.1016/j.cell.2014.10.026; pmid: 25417158
55. M. Buggert *et al*., The identity of human tissue-emigrant CD8$^+$ T cells. *Cell* **183**, 1946–1961.e15 (2020). doi: 10.1016/j.cell.2020.11.019; pmid: 33306960
56. B. V. Kumar *et al*., Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep.* **20**, 2921–2934 (2017). doi: 10.1016/j.celrep.2017.08.078; pmid: 28930685
57. N. J. Schuldt, B. A. Binstadt, Dual TCR T cells: Identity crisis or multitaskers? *J. Immunol.* **202**, 637–644 (2019). doi: 10.4049/jimmunol.1800904; pmid: 30670579
58. M. P. Soares, I. Hamza, Macrophages and iron metabolism. *Immunity* **44**, 492–504 (2016). doi: 10.1016/j.immuni.2016.02.016; pmid: 26982356

59. T. Ito *et al*., TSLP-activated dendritic cells induce an inflammatory T helper type 2 cell response through OX40 ligand. *J. Exp. Med.* **202**, 1213–1223 (2005). doi: 10.1084/jem.20051135; pmid: 16275760
60. V. Soumelis *et al*., Human epithelial cells trigger dendritic cell–mediated allergic inflammation by producing TSLP. *Nat. Immunol.* **3**, 673–680 (2002). doi: 10.1038/ni805; pmid: 12055625
61. H. J. Sutton *et al*., Atypical B cells are part of an alternative lineage of B cells that participates in responses to vaccination and infection in humans. *Cell Rep.* **34**, 108684 (2021). doi: 10.1016/j.celrep.2020.108684; pmid: 33567273
62. D. Nehar-Belaid *et al*., Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat. Immunol.* **21**, 1094–1106 (2020). doi: 10.1038/s41590-020-0743-0; pmid: 32747814
63. A. Bonner, A. Almogren, P. B. Furtado, M. A. Kerr, S. J. Perkins, The nonplanar secretory IgA2 and near planar secretory IgA1 solution structures rationalize their different mucosal immune responses. *J. Biol. Chem.* **284**, 5077–5087 (2009). doi: 10.1074/jbc.M807529200; pmid: 19109255
64. M. Kilian, J. Reinholdt, H. Lomholt, K. Poulsen, E. V. Frandsen, Biological significance of IgA1 proteases in bacterial colonization and pathogenesis: Critical evaluation of experimental evidence. *APMIS* **104**, 321–338 (1996). doi: 10.1111/j.1699-0463.1996.tb00724.x; pmid: 8703438
65. Y. Zhao *et al*., Clonal expansion and activation of tissue-resident memory-like T_H17 cells expressing GM-CSF in the lungs of patients with severe COVID-19. *Sci. Immunol.* **6**, eabf6692 (2021). doi: 10.1126/sciimmunol.abf6692; pmid: 33622974
66. E. J. Kunkel *et al*., Lymphocyte CC chemokine receptor 9 and epithelial thymus-expressed chemokine (TECK) expression distinguish the small intestinal immune compartment: Epithelial expression of tissue-specific chemokines as an organizing principle in regional immunity. *J. Exp. Med.* **192**, 761–768 (2000). doi: 10.1084/jem.192.5.761; pmid: 10974041
67. N. J. Bernstein *et al*., Solo: Doublet identification in single-cell RNA-seq via semi-supervised deep learning. *Cell Syst.* **11**, 95–101.e5 (2020). doi: 10.1016/j.cels.2020.05.010; pmid: 32592658
68. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019). doi: 10.1016/j.cels.2018.11.005; pmid: 30954476
69. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018). doi: 10.1186/s13059-017-1382-0; pmid: 29409532
70. K. Polański *et al*., BBKNN: Fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020). pmid: 31400197
71. R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018). doi: 10.1038/s41592-018-0229-2; pmid: 30504886
72. M. Büttner, Z. Miao, F. A. Wolf, S. A. Teichmann, F. J. Theis, A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods* **16**, 43–49 (2019). doi: 10.1038/s41592-018-0254-1; pmid: 30573817
73. G. Sturm *et al*., Scirpy: A Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics* **36**, 4817–4818 (2020). doi: 10.1093/bioinformatics/btaa611; pmid: 32614448
74. C. Domínguez Conde, Teichlab/TissueImmuneCellAtlas, Zenodo (2022); https://zenodo.org/record/6334988.

ACKNOWLEDGMENTS

We thank the deceased organ donors, donor families, the extended Cambridge Biorepository for Translational Medicine (CBTM) team, and the transplant coordinators at LiveOnNY for access to the tissue samples. We thank the CellTypist Annotation Team [composed of S. Webb, L. Jardine (M. Haniffa laboratory), B. Stewart, Z.K.T. (M.R.C. laboratory), R. Hoo (R. Vento-Tormo laboratory), K. James, J.P., E. Madissoon, W. Sungnak, R.E., and P. He (S.A.T. laboratory) for contributions to the harmonization of CellTypist labels. We thank V. Svensson for the initial idea of CellTypist. CellTypist is publicly available, with user-friendly documentation for application to any single-cell transcriptomics datasets (https://github.com/Teichlab/celltypist#interactive-tutorials). We thank A. Maartens for providing feedback on the text. We acknowledge C. Usher, J. Eliasova, and BioRender for graphical images. We thank M. Anderson, J. Gardner, C. Benoist, and M. Haniffa for discussions on data interpretation, and we thank C. Talavera-López and N. Kumasaka for discussions on data analysis. We also acknowledge the support received from the Cellular Generation and Phenotyping (CGaP) core facility, the Cellular Genetics Informatics team, and Core DNA Pipelines at the Wellcome Sanger Institute and the Cambridge NIHR BRC Cell Phenotyping Hub (Department of Medicine, University of Cambridge). This publication is part of the Human Cell Atlas - www.humancellatlas.org/publications. **Funding:** This work was funded, in part, by the Wellcome Trust (grants 206194 and 108413/A/15/D to S.A.T., grant WSSS 211276/D/18/Z to the CBTM, and grant 105924/Z/14/Z; RG79413 to J.L.J.), Chan Zuckerberg Initiative Seed Network (CZIF2019-002452) (to D.L.F., P.A.Si., J.L.J., and N.Y.), and NIH grants HL145547, AI128949, and AI06697 (to D.L.F. and P.A.Si.). This work was also supported by funding from the European Research Council (grant 646794 ThDEFINE to S.A.T.) and by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The CBTM is also funded by the Chan Zuckerberg Initiative. For the purpose of open access, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. The views expressed here are those of the author(s) and not necessarily those of the NIHR, Department of Health and Social Care, or National Institutes of Health (NIH). T.G. was supported by the European Union's H2020 research and innovation program "ENLIGHT-TEN" under Marie Skłodowska-Curie grant 675395. **Author contributions:** N.Y., P.A.Si., D.L.F., K.S.-P., J.L.J., and S.A.T. designed the study. K.T.M. and K.S.-P. obtained tissues from deceased organ donors. D.K.M. and E.J.N. provided premortem blood samples. C.D.C., L.B.J., D.B.R., S.B.W., S.K.H., O.S., L.M., P.A.Sz., L.R., L.B., E.S.F., H.S.M., J.P., and D.C. performed tissue dissociation and prepared single-cell sequencing libraries. C.D.C., C.X., K.P., H.W.K., Z.K.T., R.E., and E.R. analyzed and interpreted the data. C.X. and T.G. developed CellTypist. C.X., T.G., and N.H. collected CellTypist training datasets. M.P. provided support with the CellTypist package and website. L.B.J., D.B.R., S.K.H., L.T., S.P., and N.R. performed validation experiments. L.C. provided help with histological analysis. T.L. provided help with smFISH image analysis. L.K.J., O.A.B., K.B.M., M.R.C., N.Y., P.A.Si., D.L.F., J.L.J., and S.A.T. supervised experiments and data analysis. C.D.C., C.X., L.B.J., S.A.T., and J.L.J. wrote the manuscript with contributions from D.L.F., M.R.C., D.B.R., and T.G. All authors read, provided input on, and approved the manuscript. **Competing interests:** In the past 3 years, S.A.T. has worked as a consultant for Genentech, Roche, and Transition Bio and is a remunerated member of the scientific advisory boards of Qiagen, GlaxoSmithKline, and Foresite Labs and an equity holder of Transition Bio. J.L.J. reports receiving consultancy fees and grant support from Sanofi Genzyme. All other authors have no competing interests. **Data and materials availability:** Raw single-cell sequencing data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-11536. Processed data can be downloaded and interactively explored at https://www.tissueimmunecellatlas.org. Code used for data analysis can be found on Zenodo (*74*) and https://github.com/Teichlab/TissueImmuneCellAtlas.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl5197
Materials and Methods
Supplementary Text
Figs. S1 to S26
Tables S1 to S3
References (*75*–*94*)
MDAR Reproducibility Checklist

Submitted 19 July 2021; accepted 22 March 2022
10.1126/science.abl5197

Science
JOURNALS AAAS

Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**

RESEARCH ARTICLE SUMMARY

MICROBIOTA

Noninvasive assessment of gut function using transcriptional recording sentinel cells

Florian Schmidt†, Jakob Zimmermann†, Tanmay Tanna†, Rick Farouni, Tyrrell Conway, Andrew J. Macpherson*, Randall J. Platt*

INTRODUCTION: Bacteria within the gut continuously adapt their gene expression to environmental conditions that are associated with diet, health, and disease. Noninvasive measurements of bacterial gene expression patterns throughout the intestine are important to understand in vivo microbiota physiology and pathophysiology. Current methods do not offer sufficient information about transient or proximal events within the intestine without using indirect or invasive approaches that disturb normal physiology and are inapplicable to clinical practice.

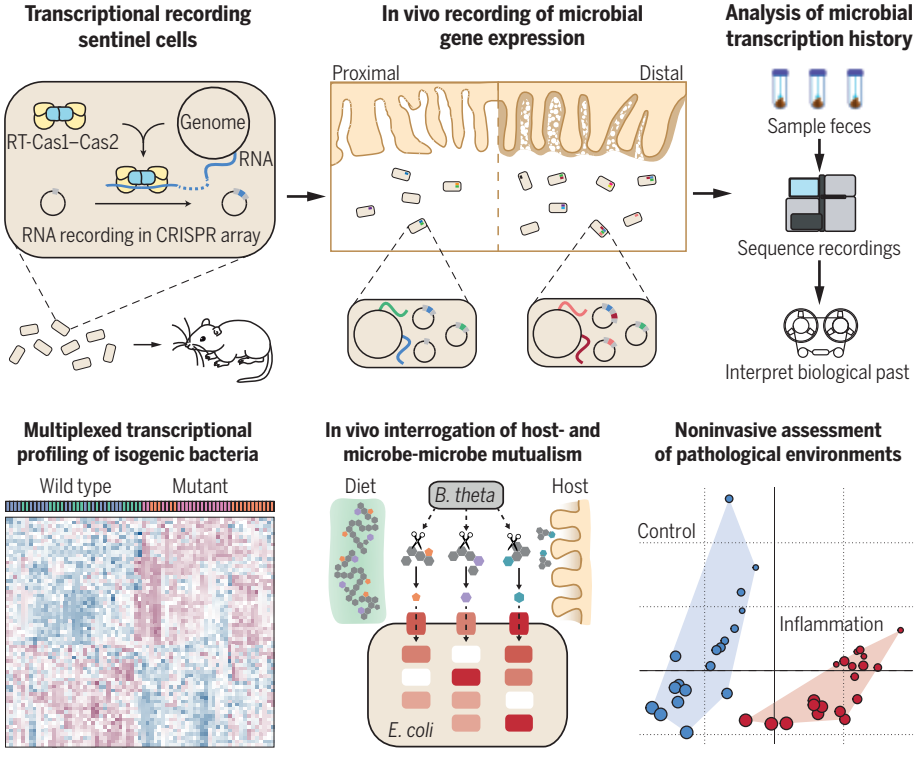
RATIONALE: Transcriptional recording by CRISPR spacer acquisition from RNA (Record-seq) enables engineered bacteria to continuously record the history of gene expression in a population of bacteria. Over time, snippets of intracellular RNA are converted into DNA and integrated as a historical record of spacer sequences within CRISPR arrays through the action of an integration complex that contains a reverse transcriptase Cas1 fusion protein (RT-Cas1) and Cas2. Here, using a refined Record-seq methodology, we used transcriptional recording *Escherichia coli* sentinel cells to reveal intestinal and microbiota physiology under different dietary and disease contexts along the length of the unmanipulated mouse intestine.

RESULTS: We used transcriptional recording sentinel cell technology in the gastrointestinal tract of germ-free and gnotobiotic mice to assess how the DNA record of spacer sequences in fecal samples uncovered distinct transcriptional records during passage from the proximal to the distal intestine, which depended on diet, inflammation, and microbe-microbe interactions.

Sentinel cells retrieved from feces of stably colonized germ-free mice accumulated new spacers over time and during intestinal transit. Fecal Record-seq profiles were distinct for mice on chow diet versus starch or fat diets. Transcriptional records of these diets were preserved in fecal spacer sequences even 2 weeks after a dietary switch, whereas diet-specific fecal bacterial RNA-sequencing (RNA-seq) profiles were rapidly lost. Direct measurements showed that Record-seq efficiently captured proximal transient transcriptional events. This included evidence of different carbon source preferences and bacterial hexuronate metabolism through the Entner-Doudoroff pathway under conditions of restricted carbon source availability, which was then verified through competitive colonization of wild-type (WT) *E. coli* and a Δ*idnK*/Δ*gntK* mutant defective in hexuronate catabolism. In addition to information about carbon source preference and metabolism, the transcriptome-scale records also provided evidence of an acid-stress response that was associated with a lowered pH in the cecum of starch-fed mice. In a dextran sulfate sodium (DSS) colitis model, sentinel cells recorded transcriptional alterations consistent with reduced anaerobic metabolism, a stringent response, and increased oxidative and membrane stress. Cocolonization with *Bacteroides thetaiotaomicron* revealed likely cross-feeding of *E. coli* from records of uptake and metabolism of fiber-derived saccharides liberated by *B. thetaiotaomicron* glycoside hydrolases.

Record-seq was also able to capture diet-specific signatures in mice colonized with a 12-organism model microbiota. Moreover, by using barcoded CRISPR arrays, we could show that Record-seq can be multiplexed in several strains of the same bacterial species that cocolonize the intestine, thus elucidating the compensatory response of a single-gene mutant to competition with the WT strain.

CONCLUSION: Transcriptional recording sentinel cells function in vivo in the mouse intestine and record transcriptome-scale information about diet, disease, and microbial interactions integrated along the length of the intestinal tract over time. Transcriptional recording enables noninvasive measurement of the intestinal tract with potential for biomedical research and future biomedical diagnostic applications. ■



Transcriptional recording sentinel cells noninvasively report interactions with diet, host, other microbes, and pathological environments. Throughout intestinal transit, sentinel cells capture information about transient mRNA expression into plasmid-encoded CRISPR arrays through the action of a reverse transcriptase Cas1-Cas2 complex. This information is retrieved by means of fecal sampling and deep sequencing followed by computational analyses. Barcoded CRISPR arrays enable transcriptional profiling of isogenic bacteria coinhabiting the intestine.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: andrew.macpherson@dbmr.unibe.ch (A.J.M.); rplatt@ethz.ch (R.J.P)
†These authors contributed equally to this work.
Cite this article as F. Schmidt *et al.*, *Science* 376, eabm6038 (2022). DOI: 10.1126/science.abm6038

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm6038

RESEARCH ARTICLE

MICROBIOTA

Noninvasive assessment of gut function using transcriptional recording sentinel cells

Florian Schmidt[1]†, Jakob Zimmermann[2]†, Tanmay Tanna[1,3]†, Rick Farouni[1], Tyrrell Conway[4], Andrew J. Macpherson[2,5]*, Randall J. Platt[1,5,6]*

Transcriptional recording by CRISPR spacer acquisition from RNA endows engineered *Escherichia coli* with synthetic memory, which through Record-seq reveals transcriptome-scale records. Microbial sentinels that traverse the gastrointestinal tract capture a wide range of genes and pathways that describe interactions with the host, including quantitative shifts in the molecular environment that result from alterations in the host diet, induced inflammation, and microbiome complexity. We demonstrate multiplexed recording using barcoded CRISPR arrays, enabling the reconstruction of transcriptional histories of isogenic bacterial strains in vivo. Record-seq therefore provides a scalable, noninvasive platform for interrogating intestinal and microbial physiology throughout the length of the intestine without manipulations to host physiology and can determine how single microbial genetic differences alter the way in which the microbe adapts to the host intestinal environment.

Engineered microorganisms that harbor molecular recording technologies enable the conversion of cellular histories into heritable nucleotide sequence archives encoded in DNA. Current recording tools—based on recombinases, single-stranded DNA recombineering, CRISPR-Cas9, and CRISPR spacer acquisition—are emerging as valuable platforms for encoding diverse biological features into DNA, including environmental metabolite concentration, cell lineage, gene expression, and horizontal gene transfer (*1*–*7*). These technologies can illuminate complex dynamic cellular processes but have largely only been deployed in vitro. An unmet need is to leverage molecular recording to engineer sentinel cells capable of traversing the gastrointestinal tract and report on microbial and host physiology in the otherwise inaccessible intestinal lumen. Current sentinel cell technologies use biosensors to report on single biomolecules (*8*–*13*) but fall considerably short of capturing the complexity of the mammalian gastrointestinal tract.

The contents of the mammalian intestine are exposed to and shape dramatically different luminal environments during transit and colonization (*14*). The microorganisms that inhabit the small intestine and colon adapt their gene expression to environmental changes associated with nutrients and pathological states. Metatranscriptomics from fecal bacteria are largely uninformative about adaptations to the proximal luminal environment (*15*), sampling of which requires invasive or disruptive procedures or use of devices that cannot preserve transient signals (*16*). Because metabolites and RNA are short-lived, omics-based measurements of transient stimuli only yield a snapshot of highly dynamic processes. An integrated measure of intestinal function therefore requires a noninvasive system that is able to sample a wide range of conditions and to preserve proximal transient signals in fecal samples.

Our overriding goal in this work was to use molecular recording to establish a scalable, noninvasive sentinel cell system for assessing intestinal and microbial physiology. To achieve this, we used Record-seq (*17*), which uses the CRISPR spacer adaptation complex of *Fusicatenibacter saccharivorans* (*Fs*RT-Cas1–Cas2) to acquire snippets of cellular RNAs as DNA within CRISPR arrays that can later be sequenced to reveal memory of past microbial gene expression. Record-seq captures transcriptome-scale information about the intestinal environment, recording the differences in microbia-host interactions according to diet, disease, or alterations of microbiota composition.

Results

Transcriptional recording sentinel cells acquire transcriptional records within the mouse gut

To establish transcriptional recording as a noninvasive recording tool in the mouse intestine, we orally gavaged germ-free C57BL/6(J) mice with *Escherichia coli* MG1655 carrying an anhydrotetracycline (aTc)–inducible transcriptional recording plasmid (Fig. 1A). Comparisons between Record-seq on sequentially collected fecal samples and contents of the ileum, cecum, and colon revealed that sentinel cells acquired increasing transcriptional records (*E. coli* genome-aligning spacers) throughout time and position along the intestinal tract (Fig. 1, B and C, and fig. S1, A and B) and according to aTc concentration (fig. S1C). The functional characteristics of *Fs*RT-Cas1–Cas2 (fig. S1, D to H) were consistent with our previous in vitro experiments (*17*, *18*).

Transcriptome-scale recording of complex and dynamic intraluminal environments

We next assessed the capacity of transcriptional recording sentinel cells to record differences in the intestinal environment by varying the animals' diet. Mice monocolonized with sentinel cells were fed one of three diets: a standard chow or a purified diet based on either starch or fat (referred to as starch or fat diets below). Starting from day 7, all groups received the chow diet (Fig. 1D and table S1). RNA-sequencing (RNA-seq) and Record-seq of fecal samples enabled the characterization of transcriptional changes and the stability of the recorded information. Before the diet switch, both RNA-seq and Record-seq readily distinguished the diet groups (fig. S2, A and B). After the switch to chow, the transcriptional signatures corresponding to starch or fat diets were rapidly lost with RNA-seq but not Record-seq (Fig. 1, E and F, and fig. S2, C to E). Thus, RNA-seq represents a snapshot measurement, whereas Record-seq durably reveals past transcriptional information.

To verify the reproducibility of Record-seq, we replicated the first 14 days in an independent experiment, largely confirming our previous observations (fig. S3, A to H). We also directly compared the two experiments and found a 95% overlap in the regulation of Record-seq differentially expressed genes (DEGs) (fig. S3I and tables S2 and S3). Thus, transcriptional recording sentinel cells reproducibly record cellular transcriptomes and provide an archive of complex and dynamic features of different mammalian intestinal environments in vivo.

Record-seq reveals adaptation of E. coli to intraluminal conditions

Record-seq characterization of the genes and pathways altered in its response to different diets delivered a detailed picture of *E. coli*'s adaptation to intraluminal environments. DEGs (table S2) readily classified the three diet conditions upon hierarchical clustering (Fig. 2A and table S4), and pairwise comparisons between each diet yielded hundreds of DEGs (Fig. 2B and fig. S4, A and B).

[1]Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel, Switzerland. [2]Department of Visceral Surgery and Medicine, Inselspital, Bern University Hospital, University of Bern, Switzerland. [3]Department of Computer Science, ETH Zurich, Universitätstrasse 6, 8092 Zurich, Switzerland. [4]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA. [5]Botnar Research Center for Child Health, Mattenstrasse 24a, 4058 Basel, Switzerland. [6]Department of Chemistry, University of Basel, Petersplatz 1, 4003 Basel, Switzerland.
*Corresponding author. Email: andrew.macpherson@dbmr.unibe.ch (A.J.M.); rplatt@ethz.ch (R.J.P)
†These authors contributed equally to this work.

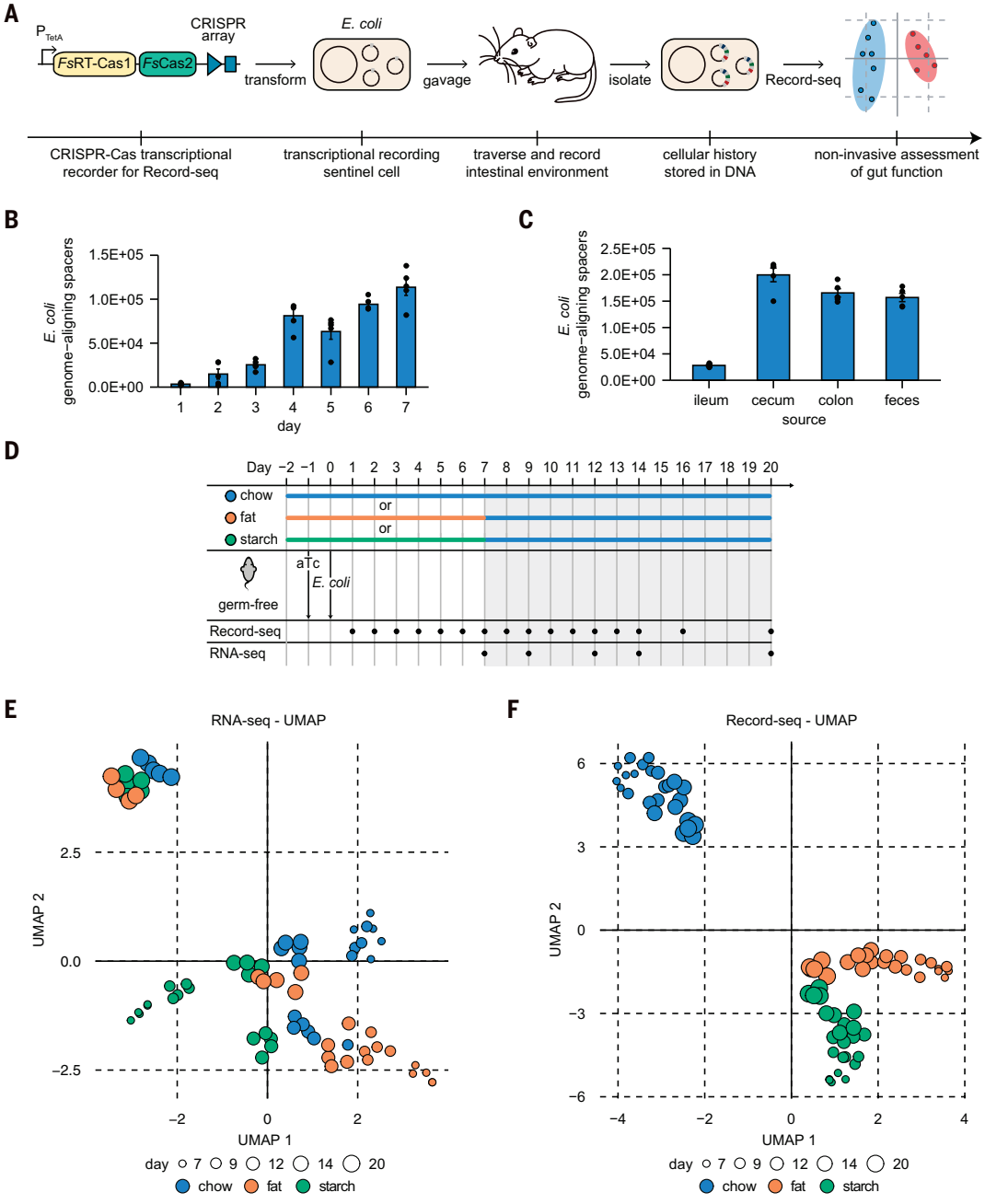


Fig. 1. Transcriptional recording sentinel cells acquire transcriptional records within the mouse gut and preserve this information throughout time. (**A**) Schematic of experimental workflow for in vivo experiments with transcriptional recording sentinel cells. *E. coli* with recording plasmid encoding aTc-inducible *Fs*RT-Cas1–Cas2 and a CRISPR array were orally gavaged into mice. Record-seq was performed on feces or intestinal contents. (**B** and **C**) Numbers of *E. coli* genome-aligning spacers (B) obtained from feces on the indicated days and (C) from the indicated intestinal sections on day 20 after gavage. (B) and (C) show mean ± SEM of $n = 5$ independent biological replicates. (**D**) Timeline of longitudinal in vivo recording experiment assessing the impact of diet on the intestinal *E. coli* transcriptome. Germ-free mice were supplied with aTc in the drinking water and gavaged with *E. coli* MG1655 transformed with the recording plasmid. Mice were fed a chow, fat, or starch diet starting 2 days before gavage until day 7. After day 7, all groups received the chow diet. Fecal RNA/Record-seq sampling is indicated, with day 7 samples being collected before changing the diets. Data for days 1 to 6, 8, 10, 11, 13, and 16 are shown in (B) and figs. S1A and S2C. (**E** and **F**) UMAP embedding of (E) RNA-seq or (F) Record-seq data from *E. coli* under chow (blue), fat (orange), or starch (green) diet on indicated days corresponding to (D). Dot sizes indicate successive time points; $n = 5$ independent biological replicates. Count thresholds were 10^4 (Record-seq) and 10^5 (RNA-seq). Outliers were excluded on the basis of modified *z*-score and relative deviation from the mean.

Pathway enrichment analysis (tables S5 and S6) (*19*) revealed a number of diet-dependent shifts in a wide range of cellular behaviors among the diet conditions (Fig. 2C and fig. S4, C and D). For further analysis, we focused on a comparison of chow versus starch diets. In starch-fed animals, we found signals indicative of nitrate use as an electron acceptor for anaerobic metabolism (*narL*), adaptation to low pH (*gadC*), potassium uptake (*kdpDE*), and carbohydrate utilization. In chow-fed animals, *E. coli* used diverse carbon sources as suggested by overrepresentation of utilization pathways for galactonate, glucarate, galactarate, sulfoquinovose, arabinose, galactose, ribose, and fucose (Fig. 2C and fig. S4E). These findings are in line with the diverse composition of the chow diet, which is enriched in plant material (table S1), as well as the previously described adaptation of intestinal symbionts to diverse polysaccharide carbon sources (*20*, *21*).

Record-seq characterization of mice on the starch diet revealed the metabolic adaptation of *E. coli* in vivo to more restricted carbon source availability. The sugar acids galacturonate and gluconate are present in the host mucus and were shown to be used as carbon sources through the Entner-Doudoroff pathway (EDP) for *E. coli* colonization in streptomycin-treated mice (*22*–*24*). In this study, we could use Record-seq directly in an unperturbed system in which the only manipulation was a diet change, revealing that gluconate and galacturonate are alternative carbon sources in the face of nutritional limitation as shown by enrichment for their catabolic pathways (Fig. 2C

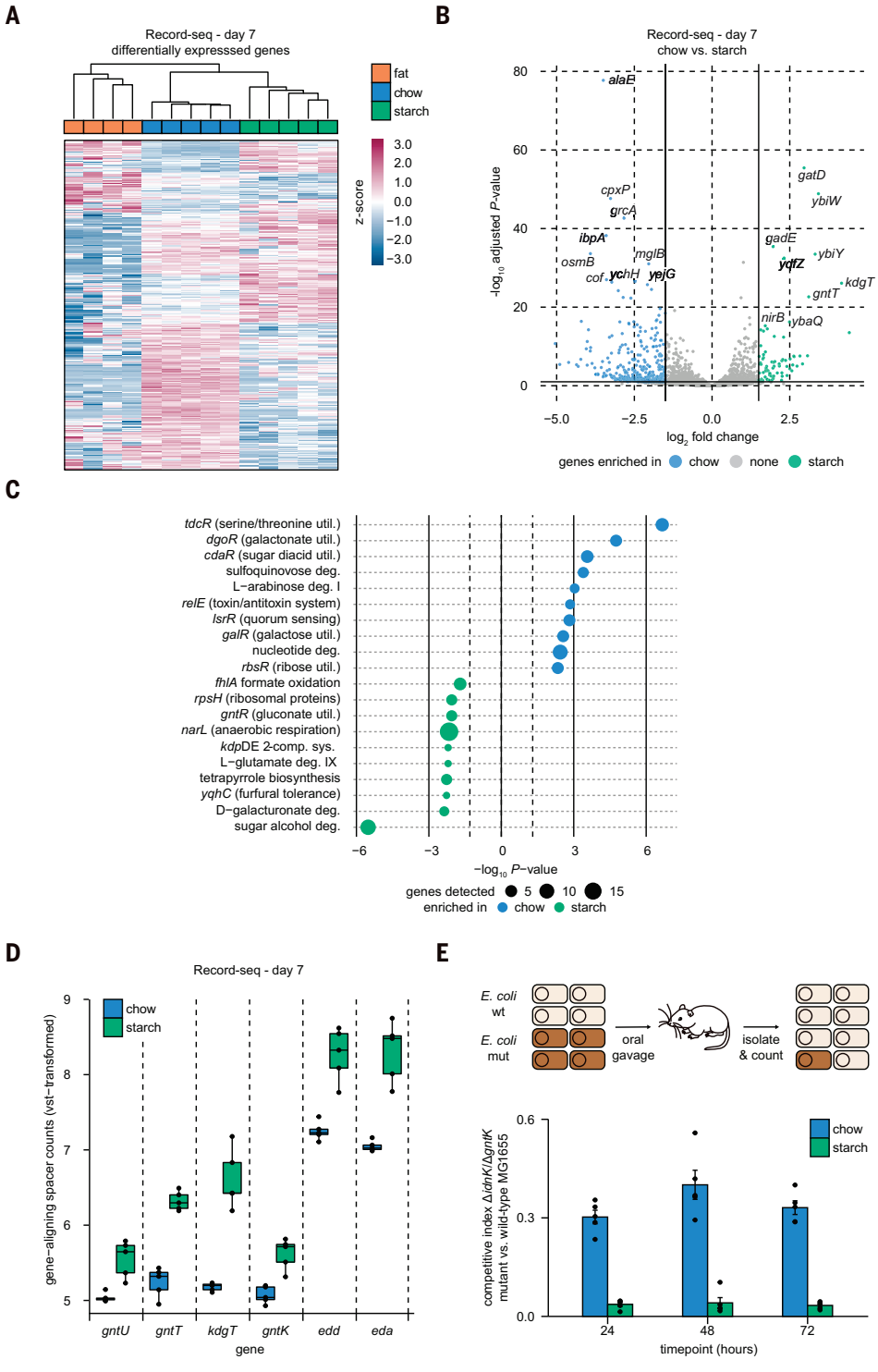


Fig. 2. Record-seq reveals transcriptional changes describing the adaptation of *E. coli* to diet-dependent intraluminal environments. (**A** to **C**) Record-seq data from day 7 feces of mice fed a chow (blue), fat (orange), or starch (green) diet. (A) Heatmap showing hierarchical clustering by using 1183 differentially expressed genes (DEGs). *z*-score standardized gene-aligning spacer counts are shown. (B) Volcano plot with DEGs (P_{adj} < 0.1, $\log_2 FC$ > 1.5) indicated. (C) Pathways and transcriptional and translational regulators enriched per diet group by use of EcoCyc. Dot sizes indicate gene numbers detected as significantly up-regulated for the respective pathway. (**D**) Box plot showing vst-transformed *E. coli* genome-aligning spacer counts for selected genes involved in gluconate metabolism corresponding to the indicated diets. (**E**) Competitive colonization experiment. Germ-free mice fed either a chow or starch diet were orally gavaged with a 1:1 mixture of WT *E. coli* MG1655 and *E. coli* MG1655 Δ*idnK*/Δ*gntK* (mut). Competitive indices were calculated from fecal recoveries as the ratio of mutant to WT CFU (mean ± SEM, n = 5 independent biological replicates, $P = 3.93 \times 10^{-17}$ likelihood ratio test representative result of two independent experiments). (A) to (D) correspond to Fig. 1D; n = 5 independent biological replicates. Count thresholds were 10^4 (Record-seq) and 10^5 (RNA-seq). Outliers were excluded on the basis of modified *z*-score and relative deviation from the mean.

and fig. S4F). These changes included key genes of gluconate metabolism, which were strongly up-regulated in *E. coli* from starch-fed mice compared with chow-fed animals, such as the low- and high-affinity gluconate transporters (*gntU* and *gntT*, respectively), the permease for 2-keto-3-deoxygluconate (*kdgT*), gluconate kinase (*gntK*), as well as the central enzymes of the EDP (*edd and eda*) (Fig. 2D).

To confirm gluconate carbon source adaptation directly, we carried out competitive colonizations with wild-type (WT) *E. coli* MG1655 and mutant *E. coli* MG1655 Δ*idnK*/Δ*gntK* (Fig. 2E). This mutant is unable to catabolize gluconate because it lacks both isoforms of the gluconate kinase, so we expected it to suffer from a substantial competitive disadvantage compared with WT cells. The competitive disadvantage of *E. coli* MG1655 Δ*idnK*/Δ*gntK* over the WT cells was nearly 10-fold greater in starch- versus chow-fed mice (Fig. 2E), suggesting that *E. coli* becomes more dependent on sugar acids from the host mucus purely because of carbon source changes. Thus, Record-seq delivers a detailed assessment of individual genes and entire pathways altered in *E. coli*'s adaptation to different diet-dependent

intraluminal environments without confounding manipulations.

Sentinel cells capture the intestinal milieu and preserve transient features of the cecum and colon

An important requirement of any intestinal reporting system that avoids potentially confounding manipulations or animal sacrifice is the ability to capture information from the microbial biomass of inaccessible proximal gut sections. Given the ability of Record-seq to preserve information over time, we directly addressed whether the sentinel cells retain information along the longitudinal axis of the gastrointestinal tract by comparing fecal Record-seq with RNA-seq of *E. coli* from different intestinal segments of mice fed a chow or starch diet. We found that ~46% of the fecal Record-seq DEGs were identified as differentially expressed in the same direction by fecal RNA-seq (Fig. 3A and table S7), highlighting concordance between the two technologies. However, ~54% of the genes detected by fecal Record-seq as up-regulated under the chow or starch diet were found to either be oppositely regulated or show no differences between the diets in fecal RNA-seq (Fig. 3A and table S7).

We considered that fecal Record-seq captures transient events in proximal sections that are absent from fecal RNA-seq. After using RNA-seq to confirm that the transcriptional states of *E. coli* from different gut segments or feces were distinct (fig. S5, A to D), we addressed the origin of the 54% of genes identified only by Record-seq in the chow condition (Fig. 3A and table S7). A substantial proportion (33%) of these genes were explained by segment-specific RNA-seq signals in the cecum and proximal colon, and a further 10% were explained by expression in the cecum and throughout the colon. Very few (1%) were expressed only in the distal colon. The remaining 56% identified with Record-seq but not RNA-seq were potentially differentially expressed in niches not covered by the sampling performed here (such as the small intestine or colonic mucus) or reflected successive transcriptional events that fell below detection thresholds in RNA-seq. Results from the starch diet were comparable with these findings, although fewer genes were selectively expressed proximally, likely owing to transit effects arising from differences between the fiber content (*25*) of the diets (Fig. 3A and table S7). Rank-based hierarchical clustering performed on genes overexpressed in the cecum compared with other sections revealed better cecal RNA-seq clustering with Record-seq than with any other RNA-seq dataset, confirming that proximal information is preserved with Record-seq (Fig. 3B and fig. S5E).

To validate a subset of insights regarding proximal gut sections revealed by fecal Record-seq, we followed up on two findings using independent methodologies. First, *uxaC*—part of the hexuronate catabolism pathway—was distinctly up-regulated in fecal Record-seq and proximal gut RNA-seq under the starch diet (Fig. 3C), suggesting that hexuronates are preferred carbon sources for *E. coli* under these conditions. In a competitive colonization between WT *E. coli* and a mutant deficient in hexuronate catabolism (Δ*uxaC*), the Δ*uxaC*

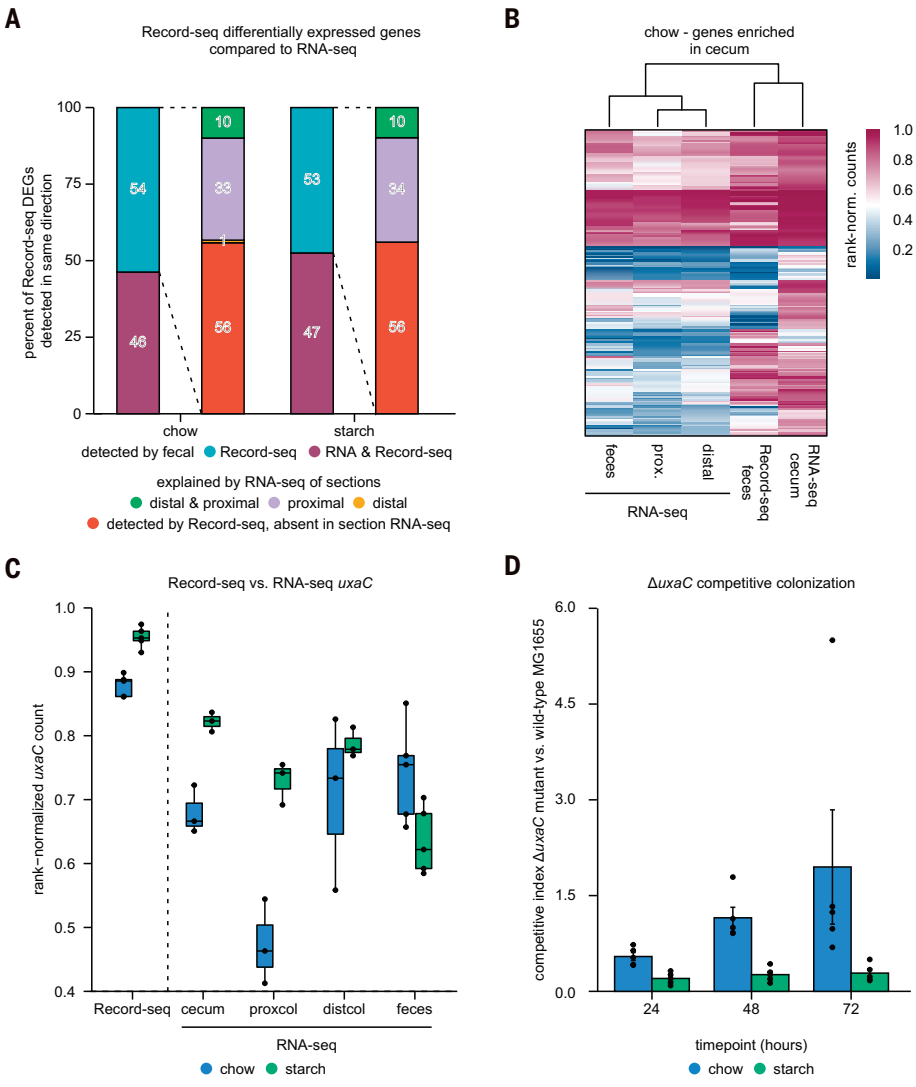


Fig. 3. Record-seq sentinel cells capture the milieu along the length of the intestine and preserve transient features of proximal large intestinal segments. Mice colonized with Record-seq sentinel cells were fed a chow or starch diet for 7 days. (**A**) Stacked bar plots showing fractions of Record-seq DEGs that were also detected with RNA-seq from feces or intestinal segments as up-regulated under a chow (430 genes) or starch diet (278 genes). Record-seq DEGs not identified with fecal RNA-seq (left bar) are categorized according to differential RNA expression in isolated gut sections (right bar): cecum or proximal colon ("proximal"), distal colon ("distal"), or both proximal and distal sections ("distal and proximal"). (**B**) Heatmap of cecum signature genes (213 genes overexpressed in the cecum) showing hierarchical clustering of rank-normalized RNA-seq and Record-seq data from the indicated intestinal sections from mice fed a chow diet. (**C**) Box plot showing *E. coli* rank-normalized counts of *uxaC* as determined with Record-seq or RNA-seq from feces, cecum, proximal colon, and distal colon corresponding to the indicated diets on day 7. Data in (A) to (C) are a combined analysis of $n = 3$ biological replicates per group, each pooled from $n = 3$ individual mice for gut sections RNA-seq, and $n = 5$ biological replicates per group for fecal RNA-seq and Record-seq. Count thresholds were 10^4 (Record-seq) and 10^5 (RNA-seq). Outliers were excluded on the basis of modified *z*-score and relative deviation from the mean. (**D**) Competitive colonization experiment. Germ-free mice fed either a chow or starch diet were orally gavaged with a 1:1 mixture of WT *E. coli* MG1655 and *E. coli* MG1655 Δ*uxaC*. Competitive indices were calculated from fecal ratios of mutant to WT CFU [mean ± SEM, $n = 5$ independent biological replicates, $P = 0.0022$ likelihood ratio test (representative result of two independent experiments)].

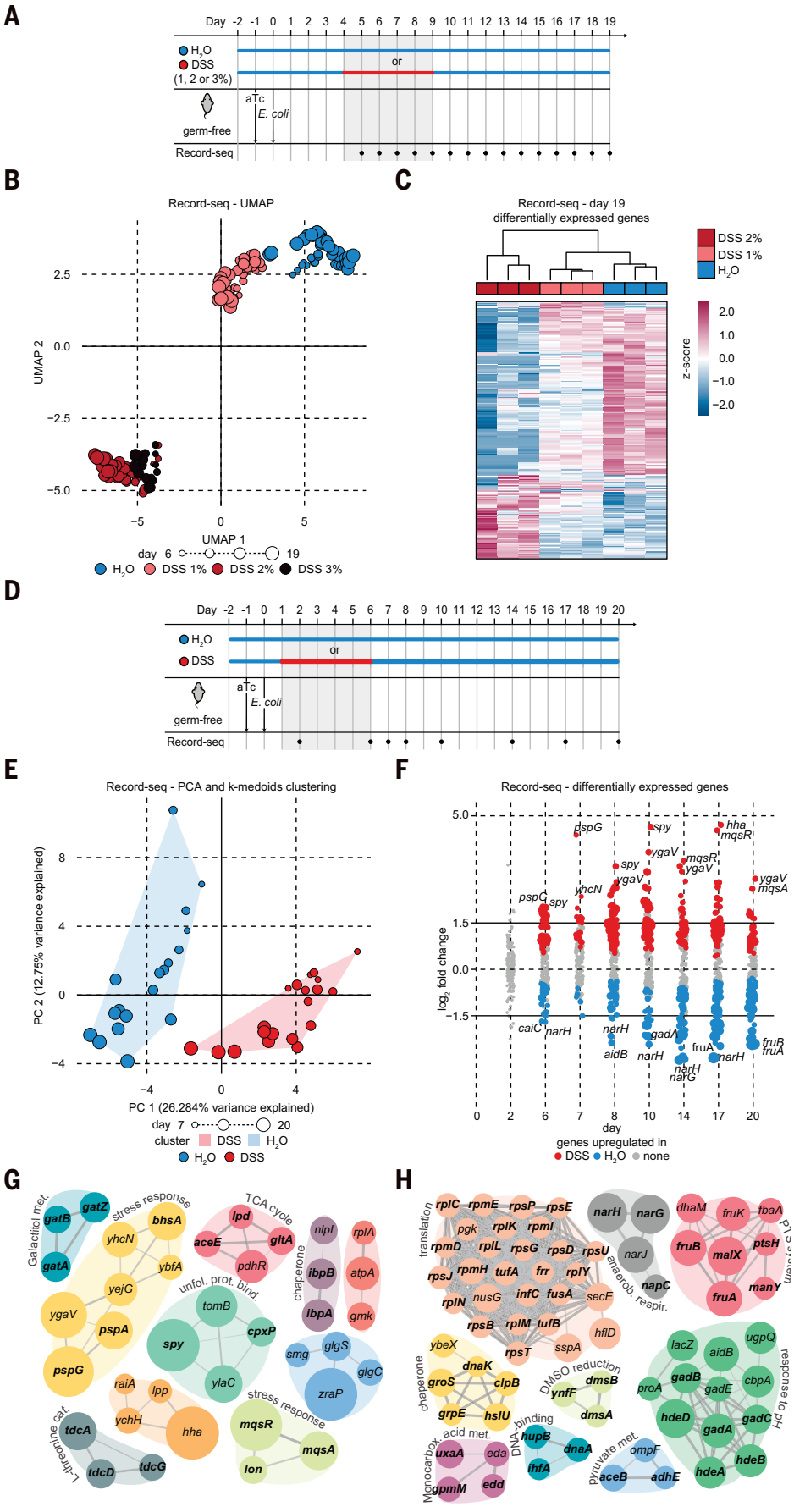


Fig. 4. Record-seq provides a non-invasive assessment of DSS-induced intestinal inflammation. (**A**) Timeline of DSS colitis recording experiment. Germ-free mice were supplied with aTc in the drinking water; gavaged with *E. coli* BL21(DE3) sentinel cells; and received 1, 2, or 3% DSS or water as indicated. Fecal Record-seq sampling is indicated. (**B**) UMAP embedding of Record-seq data for control mice (blue) or mice treated with 1% (salmon), 2% (red), or 3% (black) DSS. Dot sizes indicate successive time points. Mice receiving 3% DSS had to be euthanized on day 13. (**C**) Heatmap showing hierarchical clustering of Record-seq DEGs from control mice (blue) or mice treated with 1% (salmon) or 2% (red) DSS, day 19. *z*-score standardized gene-aligning spacer counts are shown. (**D**) Timeline of DSS colitis recording experiment. Germ-free mice were supplied with aTc in the drinking water, gavaged with *E. coli* MG1655 sentinel cells, and given 2% DSS or water as indicated. Fecal Record-seq sampling is indicated. (**E**) PCA-projected Record-seq data on days 7, 8, 10, 14, 17, and 20 for control mice (blue) or mice treated with 2% DSS (red). K-medoids clusters are indicated with convex hulls. Dot sizes indicate successive time points. (**F**) Dot plot showing $\log_2 FC$ for Record-seq DEGs identified for control mice (blue) or mice treated with 2% DSS (red). Dot sizes increase with significance (P_{adj} range, 9.8×10^{-19} to 1.0). (**G** and **H**) STRING analysis of DEGs significantly (G) up-regulated or (H) down-regulated under DSS treatment compared with that of control. Node size increases with $\log_2 FC$ [(G) 0.5 to 4.7; (H) 0.5 to 2.9]. (B) and (C) correspond to the experiment in (A), with $n = 3$ independent biological replicates. (E) to (H) correspond to the experiment in (D), with $n = 3$ to 4 independent biological replicates of each condition. Count thresholds were 5×10^3 in (B) and (C) and 10^4 in (E) to (H). Outliers were excluded on the basis of modified *z*-score and relative deviation from the mean.

strain exhibited a significantly greater competitive disadvantage in starch-fed compared with chow-fed mice (Fig. 3D). Second, acid-stress response genes (*gadABC* and *hdeAB*) were selectively overexpressed in fecal Record-seq and proximal gut RNA-seq under the starch diet (fig. S5F). Accordingly, the pH of the cecum under the starch diet was significantly reduced compared with that under the chow diet (fig. S5G). Thus, fecal Record-seq reports transcriptional events throughout the length of the intestinal tract and contains information absent from fecal RNA-seq analysis.

Record-seq provides a noninvasive assessment of intestinal inflammation

Another potential application of sentinel cells is their use as noninvasive living diagnostics capable of reporting on gastrointestinal diseases (*10*). To test this concept, we used the dextran sulfate sodium (DSS)–induced colitis mouse model. After confirming that DSS had negligible direct impact on the *E. coli* transcriptome in vitro (fig. S6A), we performed Record-seq on sequentially collected fecal samples from mice treated with increasing concentrations of DSS (Fig. 4A). Record-seq not only distinguished treated versus control mice during DSS exposure and withdrawal but could also accurately report on the phenotypic severity of the colitis model (Fig. 4, B and C, and fig. S6, B to E).

Using a longer time course and only the 2% DSS condition, we repeated the in vivo longitudinal recording experiment (Fig. 4D and fig. S6F) to characterize the intraluminal changes that result from DSS-induced inflammation throughout time. We first confirmed that transcriptional records could distinguish DSS-treated and control mice and identify DEGs throughout the duration of the experiment (Fig. 4, E and F; fig. S6, G and H; and tables S4 and S8). Next, we analyzed the DEGs with EcoCyc pathway enrichment (table S9), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) (table S10), and Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) network tools during and after inflammation and determined that genes required for bacterial membrane integrity (*pspG*) and post-stress persister cell formation (*mqsAR* and *hha*) (*26*) were up-regulated (Fig. 4, F to H; fig. S6I; and table S8). We also observed the up-regulation of chaperones (*spy* and *cpxP*) and heat-shock proteins (*ibpA* and *ibpB*) that are induced by oxidative stress (*27*, *28*). By contrast, genes and pathways most prominently down-regulated in DSS-induced colitis included those required for the nitrate and dimethyl sulfoxide (DMSO) electron acceptor pathways (*narGH* and *dmsAB*), adaptation to low pH (*gad*), and the EDP (*edd* and *eda*). Down-regulation of the EDP under inflamed conditions likely senses the mucus depletion that characterizes intestinal inflammation (*29*). These findings also suggest that *E. coli* shows decreased anaerobic respiration during inflammation but increased levels of envelope and oxidative stress, likely resulting from augmented luminal oxygenation (*30*). Several ribosomal protein operons were also down-regulated under DSS inflammatory conditions. Transcription of ribosomal proteins is regulated by *ppGpp* and *DksA* as part of

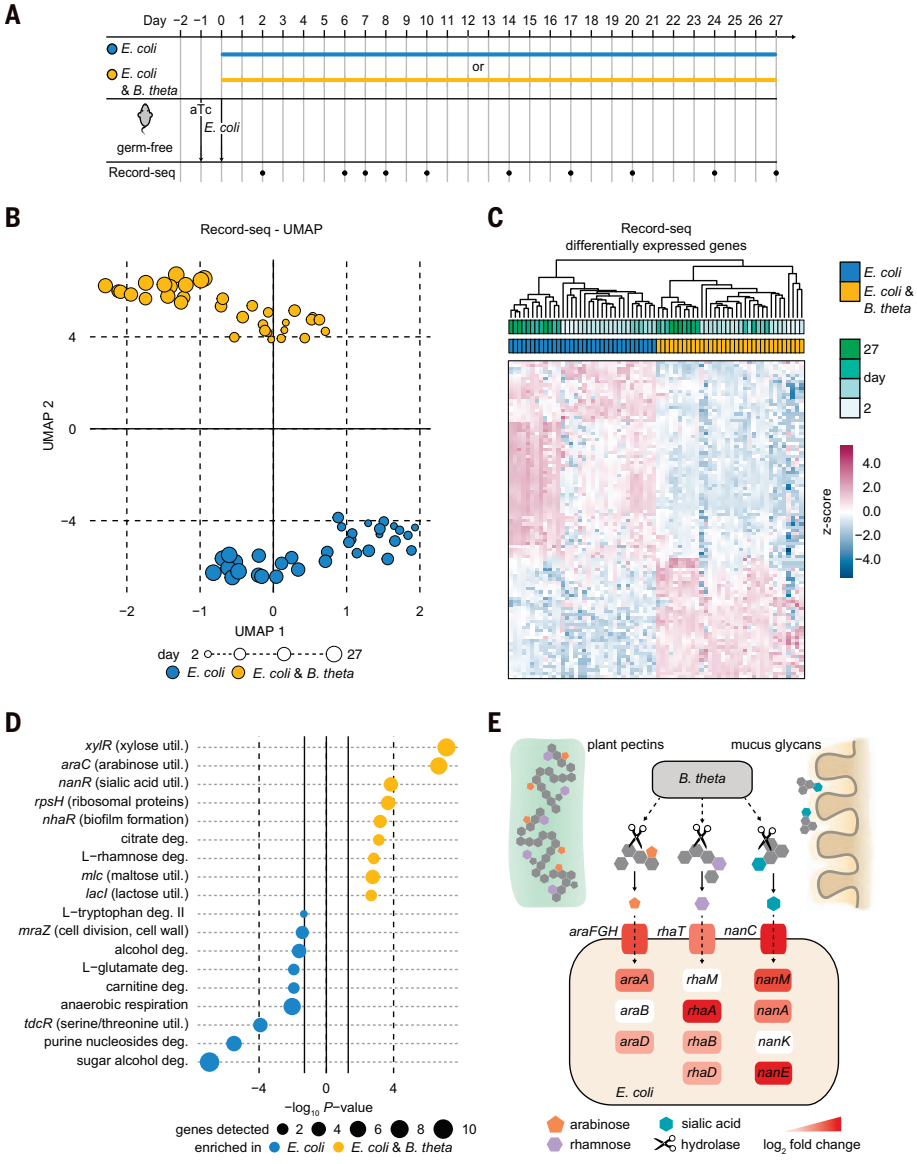


Fig. 5. Record-seq illuminates both host-microbe and microbe-microbe interactions. (**A**) Timeline of longitudinal in vivo recording experiment on interaction of *E. coli* with *B. theta* in the gut. Germ-free mice were supplied with aTc in the drinking water and gavaged with *E. coli* MG1655 sentinel cells alone or together with *B. theta*. Fecal Record-seq sampling is indicated. (**B**) UMAP embedding of Record-seq data from *E. coli* in the presence (yellow) or absence (blue) of *B. theta* on the indicated days. Dot sizes indicate successive time points. (**C**) Heatmap showing hierarchical clustering of Record-seq data from *E. coli* in the presence (yellow) or absence (blue) of *B. theta* on the indicated days by using identified DEGs. *z*-score standardized gene-aligning spacer counts are shown. (**D**) Pathways and transcriptional and translational regulators identified as enriched ($P < 0.05$) by use of EcoCyc in *E. coli* in the presence (yellow) or absence (blue) of *B. theta* on days 2 to 27. Dot sizes indicate gene numbers significantly up-regulated for the respective pathway. (**E**) Schematic depicting nutrient cross-feeding relationship between *E. coli* and *B. theta* inferred by Record-seq. *E. coli* genes encoding transporters and enzymes are depicted with color codes reflecting their Record-seq–based $\log_2 FC$ of up-regulation in the presence versus absence of *B. theta* (0 to 5.0). (B) to (E) correspond to (A), with $n = 4$ independent biological replicates of each condition. Count threshold was 5×10^3. Outliers were excluded on the basis of modified *z*-score and relative deviation from the mean.

the stringent response to nutrient limitation (*31*, *32*). Thus, Record-seq sentinel cells can accurately report on DSS-induced colitis disease severity while simultaneously revealing multiple features of the intestinal inflammatory environment.

Record-seq illuminates both host-microbe and microbe-microbe interactions

To assess the performance of Record-seq in the presence of other intestinal microbes, we started by performing longitudinal in vivo recording experiments in mice in the presence of one other prototypical member of the human gut microbiota, *Bacteroides thetaiotaomicron* (*B. theta*). Distinct transcriptional archives were obtained from mice either monocolonized with *E. coli* or cocolonized with *E. coli* and *B. theta* (Fig. 5, A to C, and tables S11 and S12). The transcriptional records revealed alterations over a wide range of microbial functions, including a dramatic shift in inferred *E. coli* carbon source preferences (Fig. 5D; fig. S7, A and B; and tables S13 and S14). In the presence of *B. theta*, pathways for utilization of xylose, arabinose, sialic acid, amino sugars citrate, rhamnose, maltose, and lactose were significantly up-regulated. Given that *B. theta* has a far richer content of polysaccharide utilization loci compared with that of *E. coli* (*21*, *33*) and that mice were fed with chow containing complex plant cell wall carbohydrates (table S1) (*25*), our data supports that cross-feeding (Fig. 5E) by *B. theta* liberates usable input nutrients (such as mono- and oligosaccharides) from otherwise unmetabolizable complex diet- or host-derived materials (such as plant pectins and mucus glycans) (*34*). Supporting the notion that nutrient cross-feeding is beneficial for *E. coli*, we observed a 3.4-fold increase in *E. coli* biomass in the presence of *B. theta* compared with that in monocolonized mice (fig. S7C).

The presence of *B. theta* also led to down-regulation of *E. coli* genes involved in metabolism of sugar alcohols, amino acids, fructose, nucleotides, and ethanolamine, which suggests that these secondary carbon sources were not required during bicolonization (Fig. 5D and fig. S7B). Selective expression of ribosomal proteins when *E. coli* cocolonized with *B. theta* (Fig. 5D and fig. S7A) is likely indicative of a stringent response on nonpreferred carbon sources when *E. coli* colonizes alone. These results, which we validated independently (fig. S7, D to G, and tables S12 to S14), demonstrate that Record-seq sentinel cells sense alterations in the intestinal environment in the presence of another taxon and report on the transcriptional adaptations that occur as a result.

We next assessed the capacity of Record-seq to characterize gut function in the presence of a complex microbiota by gavaging sentinel cells into chow- or starch-fed mice colonized by a defined 12-member sDMDMm2 consortium (Fig. 6A) (*35*). Although colonization resistance caused most gavaged *E. coli* sentinel cells to rapidly pass through the gastrointestinal tract (fig. S8A), we detected transcriptional records as early as 6 hours after gavage (fig. S8B) and differential expression according to diet from 10 hours after gavage (Fig. 6B). Additionally, the recorded information was sufficient for stratifying the diet groups (fig. S8C) and reproduced in an independent experiment (fig. S8D). By 21 hours, Record-seq detected 220 DEGs in *E. coli* passing through sDMDMm2-colonized mice on a chow versus starch diet (Fig. 6C). STRING network, KEGG/GO, and EcoCyc pathway enrichment analysis revealed diverse features of *E. coli*'s adaptation to these ecologically complex environments, including diet-dependent alterations in carbon metabolism (such as galactonate metabolism), anaerobic versus aerobic energy harvesting, and stress responses (Fig. 6D; fig. S8, E to G; and tables S15 to S18). Because none of the bacterial species from the sDMDMm2 microbiota encode a complete DeLey-Doudoroff galactonate degradation pathway, the finding of prominent up-regulation of genes involved in

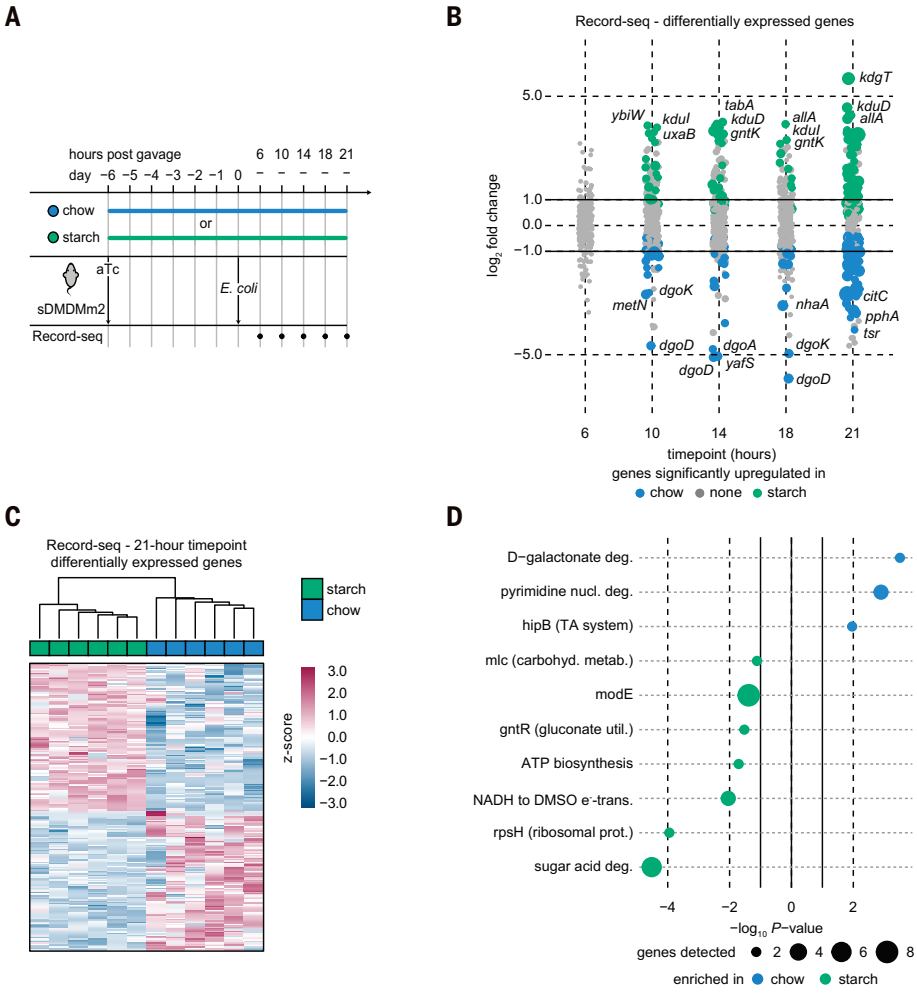


Fig. 6. Sentinel cells are deployable within a complex microbiota. (**A**) Timeline of complex microbiota recording experiment. Mice harboring a representative 12-member intestinal microbiota (sDMDMm2) were placed on either a chow or starch diet, supplied with aTc in the drinking water, and gavaged with a dose of 7×10^{10} CFU aTc-pretreated *E. coli* MG1655 as indicated. Fecal Record-seq sampling is indicated. (**B**) Dot plot showing the $\log_2 FC$ for Record-seq DEGs corresponding to the indicated diet and time. Dot sizes increase with significance (P_{adj} range, 1.9×10^{-17} to 1.0). Colored dots indicate a $P_{adj} < 0.1$, and gray dots indicate nonsignificant differences. (**C**) Heatmap showing hierarchical clustering of Record-seq data at 21 hours from mice fed a chow (blue) or starch (green) diet based on identified DEGs. *z*-score standardized gene-aligning spacer counts are shown. (**D**) Pathways and transcriptional and translational regulators identified as enriched ($P < 0.05$) in mice fed chow (blue) or starch (green) by use of EcoCyc. Dot sizes increase with number of genes detected as significantly up-regulated for the respective pathway. (B) to (D) correspond to (A), with $n = 6$ independent biological replicates for each diet. Count threshold was 5×10^3. Outliers were excluded on the basis of modified *z*-score and relative deviation from the mean.

galactonate metabolism in *E. coli* from chow-fed sDMDMm2 mice indicates that galactonate is likely available as a substrate for *E. coli* in this in vivo microbial consortium (*36*). Consistent with our earlier monocolonization results (Fig. 2C), *E. coli* from starch-fed sDMDMm2 mice overexpressed genes for utilization of mucus-derived sugar acids such as hexuronates and gluconate (fig. S8G), supporting the interpretation that host-derived sugar acids become more important even in the presence of an unmanipulated gnotobiotic microbiota as the dietary nutrient sources become less rich (*22*, *36*). Thus, transcriptional recording sentinel cells function in the context of an intestinal microbiota to reveal a portfolio of host-microbe and microbe-microbe interactions.

Multiplexed Record-seq enables parallel transcriptional profiling of isogenic bacterial strains coinhabiting the mouse intestine

Although genetic polymorphisms between taxa potentially allow RNA-seq to distinguish the transcriptional profiles of different taxa within an intestinal consortium, it is not informative of adaptive transcriptional differences between isogenic strains of the same taxon differing according to one genetic locus (*37*). We hypothesized that Record-seq could meet this need, allowing a mechanistic understanding of how a particular genetic lesion is functionally compensated within a taxon when two strains are coinhabiting the intestine.

To test this concept, we modified the Record-seq technology. First, we leveraged recent insights revealing that CRISPR spacer acquisition is aided by transcription-coupled repair (*38*) and introduced a constitutive promoter upstream of the CRISPR array within the recording plasmid, which improved recording efficiency. Second, we developed multiplexed transcriptional recording (Fig. 7A) by using two orthogonal CRISPR arrays with distinct leader and direct repeat (DR) sequences (*17*), referred to here as DR1 and DR2. After confirming that the barcoded recording constructs facilitated labeling and computational stratification of the transcriptional archives of isogenic *E. coli* strains in vitro and in vivo (figs. S9 and S10), we investigated whether multiplexed Record-seq could reveal the compensatory mechanism of an isogenic single-gene mutant in a competitive setting in vivo. We prioritized *uxaC*-deficient *E. coli* because Record-seq had revealed the importance of *uxaC* under the starch diet (Fig. 3, C and D). In two independent experiments, germ-free mice cocolonized with WT and *uxaC*-deficient (Δ*uxaC*) *E. coli* MG1655 harbored barcoded recording plasmids in which WT-DR2 was in competition with Δ*uxaC*-DR1 (group 1) or conversely WT-DR1 competed with Δ*uxaC*-DR2 (group 2) (fig. S11A). Each isogenic strain revealed robust transcriptional recording activity in vivo, and differences in the transcriptional signatures were driven by genotype (Fig. 7B and figs. S11B and S12), demonstrating multiplexed transcriptional recording of two isogenic *E. coli* strains inside the same mouse intestine.

Analysis of DEGs (table S19) and pathways (tables S20 and S21) revealed decreased expression by the *uxaC*-deficient mutant of other hexuronate utilization genes such as *uxaA*, *uxaB*, *uxuA*, and *uxuB* in addition to the expected lack of *uxaC* (Fig. 7C and fig. S12C). Because hexuronate utilization genes are induced by galacturonate and/or glucuronate (*39*), the WT strain likely displaces the Δ*uxaC* strain from niches where these sugar acids are available. Furthermore, the Δ*uxaC* strain appears to compensate for this deficiency through the up-regulation of serine/threonine– and maltose-utilizing gene products. Thus, Record-seq has the distinctive capacity to enable multiplexed transcriptional profiling of two isogenic strains of *E. coli* in the same mouse gut. This approach reveals compensatory mechanisms in response to intraspecies competition in which RNA-seq–based experiments are uninformative.

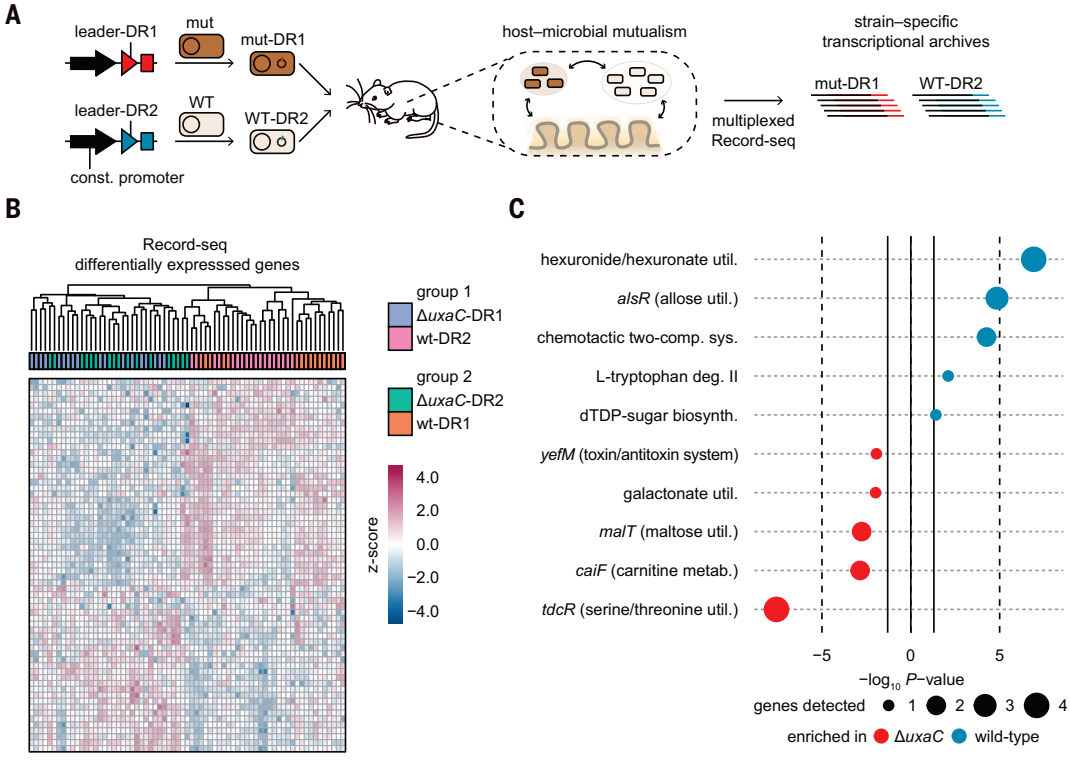


Fig. 7. Sentinel cells enable multiplexed transcriptional profiling of isogenic bacterial strains coinhabiting the mouse intestine. (**A**) Schematic for multiplexed in vivo recording with sentinel cells barcoded by their CRISPR array. Recording plasmids with either a leader-DR1 or leader-DR2 CRISPR array were transformed into WT or mutant *E. coli* cells and orally gavaged into mice at a 1:1 ratio. Spacers were assigned to the appropriate strain by using the leader-DR barcode. (**B**) Germ-free mice on a starch diet were orally gavaged with a 1:1 mixture of WT and *uxa*C-deficient barcoded *E. coli* sentinel cells. Two different pairings between genotype (WT or Δ*uxaC*) and CRISPR array (DR1 or DR2) were used. Group 1 was given Δ*uxaC*-DR1 (blue) and WT-DR2 (pink). Group 2 was given Δ*uxaC*-DR2 (green) and WT-DR1 (orange). Fecal Record-seq samples were collected daily from day 7 to day 10, and a combined heatmap is shown. Hierarchical clustering was performed by using the top 50 DEGs detected in both comparisons. *z*-score standardized gene-aligning spacer counts are shown. (**C**) Pathways and transcriptional and translational regulators identified as enriched ($P < 0.05$) by use of EcoCyc on the basis of high-confidence DEGs between Δ*uxaC* (red) and WT (blue) *E. coli*. Dot sizes indicate gene number significantly up-regulated for the respective pathway. (B) and (C) correspond to fig. S11A, with $n = 5$ independent biological replicates of each condition and days 7 to 10. Count threshold was 10^4. Outliers were excluded on the basis of modified *z*-score and relative deviation from the mean.

Conclusions

We demonstrated that transcriptional recording sentinel cells by using *Fs*RT-Cas1–Cas2 to integrate RNA-derived spacers from the *E. coli* transcriptome into plasmid DNA-encoded CRISPR arrays are capable of recording complex and dynamic transcriptional changes during the adaptation of *E. coli* throughout time, transit, and perturbation of the mammalian intestinal tract. This scalable, noninvasive system for assessing intestinal function in vivo archives characteristic microbial signatures of physiological or pathological states. Transcriptome-scale recordings elucidate microbial responses to alterations in the intraluminal environment across nutrition, intestinal inflammation, and microbe-microbe interactions. We have illustrated how carbon preferences can be shown with the tool without confounding manipulations and validated findings of intraluminal microbial adaptation under different dietary conditions.

Record-seq offers multiple advantages compared with contemporary techniques. First, unlike conventional cell-based biosensors, Record-seq does not require a specific biosensor for every biomolecule of interest and can report on a wide range of complex biological features, serving as an unbiased discovery tool. Second, compared with conventional omics-based technologies run on fecal samples, Record-seq integrates information on gut function along the length of the intestine, which is particularly valuable for studying the proximal large intestinal environment that has been largely refractory to detailed studies because of its inaccessible location. Third, multiplexed Record-seq reveals diverse in situ microbe-microbe interactions, even between variants of a single microbial species within the same animal over time, which cannot be readily scaled or implemented in high throughput with conventional methods.

We envision that transcriptional recording sentinel cells and the Record-seq principle may enable understanding intestinal and microbiota physiology under different dietary conditions, disease contexts, or constraints in humans where longitudinal sample collection based on surgical, endoscopic, fecal, ingestible device, or post mortem methods are neither feasible nor sufficient.

Materials and methods

Bacterial strains

Bacterial strains used in this study were *Escherichia coli* strains MG1655 (ATCC no. 700926) and BL21(DE3) Gold (Agilent no. 230132) and *Bacteroides thetaiotaomicron* strain VPI-5482 (ATCC no. 29148). *E. coli* MG1655 Str^R, *E. coli* MG1655 Str^R Nal^R, *E. coli* MG1655 Str^R Δ*idnK*/Δ*gntK*, and *E. coli* MG1655 Str^R Δ*uxaC* were provided by T. Conway and the Kan^R marker of the Δ*uxaC* strain was removed using pCP20 recombination as reported previously (*40*) yielding *E. coli* MG1655 Str^R Δ*uxaC* ΔKan^R. All *E. coli* strains used in this study are reported in table S22. MG1655 isolates have been sequenced and their genomic sequences are available in the NCBI Assembly database (PRJNA807125). NCBI Reference Sequences U00096.3 and NC_012947.1 were used for MG1655 and BL21(DE3) Gold, respectively. The stable defined moderately diverse mouse microbiota 2 (sDMDMm2) has been described previously (*36*, *41*) and is available through the Deutsche Sammlung für Mikroorganismen und Zellkulturen DSMZ. The constituting taxa were originally isolated from the mouse intestine and comprise *Bacteroides* I48, *Blautia* YL58, *Akkermansia* YL44, *Bacteroidales* YL27, *Ruminococcaceae* KB18, *Lactobacillus* I49, *Lachnospiraceae* YL32, *Erysipelotrichaceae* I46, *Enterococcus* KB1, *Flavonifractor* YL31, *Parasutterella* YL45 and *Bifidobacterium* YL2 (table S23).

Mice

All mouse experiments were performed in accordance with Swiss federal and cantonal regulations under permit numbers BE43/16, BE44/18 and BE107/20. Germ-free C57BL/6(J) mice were born and housed in flexible-film isolators in the Clean Mouse Facility, University of Bern, Switzerland. Unless noted otherwise, mice received a vitamin-fortified rodent chow diet (Kliba Nafag 3307) sterilized by autoclaving for 20 min at 132°C and water ad libitum. Age and sex-matched mice were used at 6 to 15 weeks of age (mostly 8 to 12 weeks). Mice were constantly and independently confirmed to be germ-free within the breeding isolators by culture-dependent methods (liquid cultures in brain-heart infusion (BHI) broth (Thermo Fisher Scientific) aerobically at 37°C at 180 rpm and anaerobically in an anaerobic cabinet (Meintrup DWS) containing 80% N_2, 10% H_2, and 10% CO_2 at 37°C without shaking) and culture-independent methods (i.e., microscopic examination of fecal smears stained with the DNA dye SYTOX green (Thermo Fisher Scientific)).

During experiments, the absence of bacteria other than *E. coli* (and *B. theta* in experiments related to Fig. 5 and fig. S7) was constantly confirmed by culturing of fecal suspensions on lysogeny broth (LB) agar aerobically and on BHI agar with 5% defibrinated sheep blood anaerobically. The microbial biomass per gram feces was determined by weighing fecal pellets, homogenizing them in 1 ml of PBS using a Retsch MM400 tissue lyser at 30 Hz for 3 min and streaking serial dilutions in PBS onto LB agar.

One day before gavage, drinking water of the mice was exchanged for water containing 30 μg/ml of anhydrotetracycline (aTc) (Adipogen) and 100 μg/ml of kanamycin sulfate (MP biochemicals) which was prepared by diluting stock solutions of 2 mg/ml of aTc in 95% ethanol and 100 μg/ml of kanamycin sulfate in aqua bidest into sterile tap water.

Plasmid transformation

For plasmid transformation, *E. coli* BL21 (DE3) Gold (Agilent Technologies) and *E. coli* MG1655 (ATCC no. 700926) were made chemically competent using the Mix & Go *E. coli* Transformation Kit & Buffer Set (Zymo Research). For this, strains were streaked to single colonies on LB (Difco) agar (Huberlab) plates without antibiotics followed by growth overnight at 37°C. A single colony of *E. coli* was inoculated into 50 ml of ZymoBroth (Zymo Research) and grown at 19°C, 220 rpm in an orbital shaker (New Brunswick Innova 40R) to an optical density of OD_{600} = 0.45. Subsequently, cells were made competent following the manufacturers protocol, dispensed to aliquots of 25 μl, flash-frozen in liquid nitrogen, and stored at −80°C.

Transformation with the recording plasmid pFS_0453 (Addgene #117006) was performed by adding ~60 ng of plasmid DNA to 25 μl of competent cells, followed by heat shock (42°C, 30 s), recovery in 120 μl of S.O.C medium at 37°C, 900 rpm, 30 min and spreading on LB agar plates containing 50 μg/ml of kanamycin sulfate (Biochemica). Glycerol stocks were created by growth of transformants in LB with 50 μg/ml of kanamycin sulfate at 37°C, 180 rpm in bacterial culture tubes followed by mixing of 500 μl of saturated culture with 500 μl of sterile filtered 50% (v/v) glycerol and freezing at −80°C for long-term storage.

Oral gavage

Unless stated otherwise, a saturating gavage dose of 1×10^9 colony forming units (CFU) of *E. coli* was used to avoid confounding the biological signal reported by our sentinel cells with an initial expansion in the gastrointestinal tract (*42*). To maintain the recording plasmid and ensure functional stability of the sentinel cells, we added kanamycin sulfate to the drinking water and confirmed that the transformed cells colonized at $(7.9 \pm 2.6) \times 10^9$ CFU/g feces, comparably to what has been reported for the parental strain (*42*). For oral gavage, *E. coli* MG1655 or BL21(DE3) each transformed with pFS_0453 was inoculated from freshly grown colonies and cultured overnight under aerobic conditions in LB containing 50 μg/ml of kanamycin sulfate at 37°C, 180 rpm. Upon saturation, cultures were centrifuged at 3480*g* for 10 min at room temperature and washed twice with the equivalent volume as the LB culture in sterile phosphate-buffered saline (PBS) (8 g per liter of NaCl, 0.2 g per liter of KCl, 1.44 g per liter of Na_2HPO_4, and 0.24 g per liter of KH_2PO_4, all from Sigma-Aldrich). The required dose of

bacteria was resuspended in PBS (1×10^9 CFU per 500 µl). The bacterial suspension was orally gavaged directly into the mouse duodenum with a 12 gauge straight stainless-steel needle (Provet AG) attached to a 2-ml syringe. The culture volume was chosen according to the number of mice to be gavaged with the equivalent of 3 ml of saturated culture being gavaged into each mouse. Culture vessels were at least twice as large as the culture volume to ensure proper aeration. For example, to gavage $n = 15$ mice, *E. coli* was grown in 100 ml of LB broth in a 250-ml bottle, washed twice with 100 ml of PBS, and resuspended in 16.6 ml of PBS from which 500 µl was gavaged into each mouse. Gavage doses and absence of contamination were confirmed by streaking serially diluted suspensions onto LB agar.

Isolation of RNA and RNA-seq

Fecal pellets were homogenized in 1 ml of PBS using a Retsch MM400 tissue lyser at 30 Hz for 3 min. Large particles were pelleted by centrifugation at 200*g* for 2 min at room temperature. Bacteria in the supernatant were pelleted by centrifugation at 6800*g* for 3 min at room temperature and lysed in 100 µl of RNA extraction solution containing 95% (v/v) formamide (VWR), 0.025% (w/v) SDS (Bio-Rad), 18 mM EDTA (Merck Millipore), and 1% 2-mercaptoethanol (Merck Millipore) (*43*) at 95°C, 1200 rpm shaking for 7 min. Following centrifugation at 16,000*g* for 5 min at room temperature, the RNA in the supernatant was purified with the RNeasy clean-up kit (QIAGEN) following the manufacturer's instructions, including the optional DNase treatment (15 min at room temperature). RNA was frozen at −80°C for storage and submitted to the Next Generation Sequencing (NGS) Platform Bern for ribosomal RNA (rRNA) depletion using the RiboMinus Transcriptome Isolation Kit, bacteria (Invitrogen), followed by library preparation using the Illumina TruSeq Stranded Total RNA Kit (Illumina) and sequencing on an Illumina NovaSeq platform using the NovaSeq 6000 SP Reagent Kit (100 cycles).

Isolation of plasmid DNA from feces and intestinal contents.

For *E. coli* monocolonized mice, 50-100 mg of fecal material or intestinal contents were collected and frozen at −20°C. After thawing, feces or intestinal contents were homogenized in 1 ml of PBS using a Retsch MM400 tissue lyser at 30 Hz for 3 min. Large particles were pelleted by centrifugation at 200*g* for 2 min at room temperature. Bacteria in the supernatant were pelleted by centrifugation at 6800*g* for 3 min at room temperature and plasmid DNA was isolated using the QIAprep spin Miniprep kit (QIAGEN) with the following modifications: *E. coli* cells were resuspended in 500 µl of buffer P1 and then lysed by the addition of 500 µl of buffer P2 for 5 min at room temperature. The reaction was neutralized by the addition of 700 µl of buffer N3 and centrifuged at 18,000*g* for 10 min at room temperature to pellet debris. Supernatant was passed through a spin column on a vacuum manifold (QIAGEN), upon which the column was washed with 500 µl of buffer PB followed by 700 µl of buffer PE. Residual wash buffer was removed by centrifugation at 18,000*g* for 1 min at room temperature. For cecal contents of monocolonized mice (~200 mg), buffer volumes were increased to 1000 µl of buffer P1, 1000 µl of buffer P2, and 1400 µl of buffer N3, whereas volumes of buffer PB and PE remained the same. Plasmid DNA was eluted from the column by the addition of 50 µl of pre-warmed buffer EB (55°C) followed by incubation at 55°C at 850 rpm for 3 min followed by centrifugation at 10,000*g* for 1 min. A total of three elution steps with 50 µl were performed yielding 150 µl of eluate. DNA from this eluate was precipitated by the addition of 15 µl of 3 M sodium acetate solution (Sigma-Aldrich) and 150 µl of 2-propanol (Merck-Millipore) followed by incubation at −20°C for 20 min and centrifugation at 20,000*g* for 20 min at room temperature. Supernatant was carefully removed without disturbing the pellet which was washed with 500 µl of 80% (v/v) ethanol and centrifuged at 20,000*g* for 15 min at room temperature. Ethanol was removed completely and the pellet was briefly dried (55°C, 30 s) before resuspension in 15 µl of buffer EB and transfer to 96-well PCR plates for storage at −20°C or immediate use in SENECA.

Quantification of isolated plasmid DNA by droplet digital PCR (ddPCR)

For quantification by ddPCR, plasmid DNA was first diluted 100,000-fold in ddPCR dilution buffer. ddPCR dilution buffer consisted of 2 ng/µl of sheared salmon sperm DNA (Sigma Aldrich) and 0.05% Pluronic F-68 (Invitrogen) in UltraPure DNase/RNase-Free distilled water (Thermo Fisher Scientific). Dilution steps were performed in twin.tec PCR plate 96 LoBind (Eppendorf) as follows: dilution 1: 1 µl of plasmid DNA, 49 µl of ddPCR dilution buffer; followed by dilution 2: 1 µl of dilution 1, 49 µl of ddPCR dilution buffer; and finally dilution 3:1 µl of dilution 2, 39 µl of ddPCR dilution buffer. Primer-probe assays targeting *Fs*RT-Cas1 were prepared by mixing of 180 µl of FS_2814 (100 µM), 180 µl FS_2815 (100 µM), and 50 µl of FS_2816 (100 µM) with 590 µl of TE buffer (Sigma Aldrich) (table S24). Aliquots of 100 µl of primer-probe assay were prepared and stored in 1.5-ml amber tubes (Eppendorf) at −20°C. ddPCR was performed by mixing 4.5 µl of dilution 3 template, 1.1 µl of primer-probe assay, 0.25 µl of FastDigest XhoI (Thermo Fisher Scientific), 11 µl of ddPCR Supermix for probes (no dUTP) (Bio-Rad), and 5.4 µl of UltraPure DNase/RNase-Free distilled water per reaction. PCR reactions were dispensed into droplets using the QX100 droplet generator (Bio-Rad) according to the manufacturer's protocol. PCR amplification was performed in an Eppendorf Mastercycler Gradient (95°C for 10 min, followed by 42 cycles of 95°C for 30 s, 57.1°C for 60 s, 72°C for a 15-s final extension, and a final 98°C 10-min step) with a ramp rate for all steps of 2°C/s as specified by the manufacturer. Readouts were measured using a QX100 droplet reader (Bio-Rad) with a cut-off for positive droplets manually set to 3500.

Selective amplification of expanded CRISPR arrays (SENECA)

The SENECA library preparation method has been extensively described before (*18*). All DNA oligonucleotides were ordered from IDT (table S25), FastDigest FaqI (ThermoFisher Scientific), T7 DNA Ligase and NEBNext High Fidelity PCR Master Mix, 2X (both New England Biolabs). Due to the low concentration of plasmid DNA extracted from fecal pellets and residual genomic DNA, input DNA was not normalized but instead 3.75 to 7.5 µl of purified plasmid DNA was used for SENECA adapter ligation, volumes are stated in the subsections below corresponding to the respective experiments along with the specific annealed oligonucleotides (carrying library barcodes) for adapter ligation. SENECA first round PCR was performed with 23 cycles instead of 22 cycles and otherwise as described before (*18*). After second-round PCR, 3 µl of each sample were mixed with 17 µl of UltraPure DNase/RNase-free distilled water (Thermo Fisher Scientific) and loaded on an E-Gel 48 Agarose Gel, 2% along 150 ng of GeneRuler low-range DNA ladder (Thermo Fisher Scientific) in 20 µl for gel-based quantification using the software Bio-Rad Image Lab version 6.0.1. Samples from an experiment were assigned to five bins based on their DNA concentrations and pooled according to these bins. Each sub-pool was purified separately by PCR purification and gel extraction from E-GelEX 2% (Thermo Fisher Scientific) agarose gels as described previously, individually quantified by quantitative PCR (qPCR) using the KAPA Library Quantification Kit for Illumina Platforms (Roche), pooled to achieve equal sequencing depth according to the number of samples in each subpool and sequenced on an Illumina NextSeq 500/550 platform as described before. As Record-seq is a population-based measurement requiring many cells to reconstruct a cellular history, and in vivo experiments present a material-limiting environment, we addressed the technical inputs and outputs of our workflow (fig. S1, A and B). Record-seq usually used an input, of approximately $(4.7 \pm 0.7) \times 10^8$ cells per sample (3 to 4 fecal pellets of ~27 mg each

per mouse with a biomass of 5.6×10^9 sentinel cells per gram). After 24 hours of in vivo recording, the Record-seq output was $(3.6 \pm 0.6) \times 10^3$ spacers, which increased by $(1.0 \pm 0.5) \times 10^4$ spacers per day. After 7 days of recording, we found $(1.2 \pm 0.1) \times 10^5$ spacers, which aligned to 91 ± 1% of the 4419 transcripts in the *E. coli* transcriptome.

Primary analysis of data

The single-end sequencing readout from Record-seq and RNA-seq was processed and analyzed using a two-stage computational pipeline as described before (*18*). The first step in the primary analysis pipeline involved pre-processing of sequencing reads using FastQC v0.11.4 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmomatic v0.35 (*44*). For Record-seq, FASTQ files containing sequencing results were converted to FASTA files using the FASTX-toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit). Reads without the library barcode were excluded and acquired unique spacer sequences were identified from the remaining reads using a dedicated python script which incorporates fuzzy string matching (spacerExtractor.py). Identified unique spacer sequences for Record-seq and sequencing reads for RNA-seq were aligned to merged references that contain *E. coli* genome and plasmid sequences and annotation using Bowtie 2 (*45*). The following *E. coli* reference genomes were used: *E. coli* str. K-12 substr. MG1655 (GenBank U00096.3) and *E. coli* BL21-Gold(DE3) pLysS AG (Ensembl ASM2366v1). Alignments were then processed and assigned to either the reference *E. coli* genome or plasmid using Samtools v1.3 (*46*), and for Record-seq, duplicate alignments were removed using a custom python script (SErmdup.py). This two-step stringent filter for duplicate spacers was incorporated to remove multiple instances of the same spacer arising due to amplification or plasmid replication, and thus obtain a conservative estimate of spacer diversity. Count matrices were generated by quantifying alignments using featureCounts from the Subread package (*47*). These count matrices contain transcript counts for each RNA-seq sample and transcript-aligning spacer counts for each Record-seq sample. All the steps of this primary analysis pipeline were implemented in Snakemake (*48*) workflows. The Snakemake workflows for each experiment described in this manuscript, python scripts, reference genomes, and ancillary scripts for primary analysis are available on GitHub (Data and materials availability).

Secondary analysis of data

Secondary analysis was performed on generated count matrices by building upon the previously described recoRdseq package implemented in R (*18*). This broadly involved unsupervised clustering of samples and identification of classifier genes based on differential expression analysis. Complete analyses for all experiments described in this manuscript are available as R notebooks on GitHub (Data and materials availability). In general, the first step involved filtering count matrices by excluding outlier samples with low cumulative counts ($C = \sum_t x_{t,s}$; x = count, t = transcript, s = sample) among replicates using an empirically adjusted absolute cumulative counts threshold, as well as a combined threshold for outliers as described below:

(i) include replicate if modified Z-score $Z_i > -3$ ($Z_i = \frac{0.675^*(y_i - \tilde{y})}{\text{median}(|y_i - \tilde{y}|)}$, where y_i = cumulative count for replicate i, $\tilde{y}$ = median cumulative count for all replicates), else

(ii) if $Z_i < -3$, include replicate if relative deviation from the mean of replicates $D_i < 0.25$ $D_i = \left|\frac{y_i - \bar{y}}{\bar{y}}\right|$, where y_i = cumulative count for replicate i, and $\bar{y}$ = mean cumulative count for all replicates).

Lowly abundant (or recorded) transcripts, defined as transcripts having a low cumulative count across samples ($\sum_s x_{t,s}$; x = count, t = transcript, s = sample), were also excluded from the analysis. Further, the first day after gavage (Day 1) generally yielded low spacer counts and noisy data compared to later days in multi-day experiments, hence Day 1 was excluded from all subsequent analyses. The count matrices were then normalized and transformed using the variance-scaling transformation (VST) implemented in the DESeq2 package (*49*), which rendered the data approximately homoscedastic. For dimensionality reduction and unsupervised cluster discovery, principal component analysis (PCA) using the R base stats package and Uniform Manifold Approximation and Projection (UMAP) (*50*) using the umap package implemented in R were performed on the vst-transformed count matrices. UMAP parameters and hyperparameters were tuned for each dataset to achieve optimal separation between experimental groups. k-medoids clustering was used to detect clusters in PCA-projected data. Fixed random number seeds were used to ensure reproducibility of clustering algorithms. PCA and UMAP results as well as other illustrative plots were plotted using the ggplot2 package in R. Differential expression analysis was performed using the Wald test (pairwise comparisons) or likelihood ratio test (multiple-group comparisons) implemented in DESeq2 and the quasi-likelihood F-test implemented in the edgeR package in R. DEGs were defined as the intersect of significant genes ($P_{adj} < 0.1$, where P_{adj} = Benjamini-Hochberg adjusted P-value detected by these two tools). For time-course datasets, DEGs were independently identified for each timepoint and combined differentially expressed gene lists, ordered by the number of independent timepoints each individual gene was detected on, were generated for downstream pathway analysis. Volcano plots were generated for individual timepoints using the $\log_2$-fold change ($\log_2FC$) and p_{adj} values calculated by DESeq2. For time-course analysis, $\log_2FC$ for all genes in the combined differentially expressed gene lists were plotted over time, with point size indicating P_{adj} values. For downstream pathway analyses, the $\log_2FC$ for each gene (and each comparison) was defined as the maximum $\log_2FC$ detected for that gene over the time-course. Hierarchical clustering was performed on vst-transformed counts of DEGs after Z-score standardization for each gene, and heatmaps were generated using the pheatmap package in R.

EcoCyc pathway enrichment analysis was performed using the Fisher Exact test for lists of DEGs generated for each experiment, and pathway enrichment plots were created using the top hits. Further, network analysis was performed using the StringApp package (*51*) in Cytoscape with a confidence score ≥0.4 for these DEGs. MCL clustering using the clusterMaker2 package (*52*) was used to generate gene clusters within gene networks detected using StringApp analysis in Cytoscape, and the functional enrichment function was used to annotate these clusters using KEGG pathways, UniProt keywords, NetworkNeighborAL, GO Process, GO Function, GO Component. Nodes contributing to functional enrichment of a cluster were marked by bold font. The size of a node corresponding to each gene in the STRING networks was adjusted to reflect the detected $\log_2FC$ value for that gene. Over-representation analysis (OA) was performed for differentially expressed gene lists from each experiment based on both the GO resource (*53*) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (*54*) using the clusterProfiler package (*55*) in R. Wherever experiments were replicated, regression analysis was performed to examine the similarity of regulation in the latter experiment for DEGs detected in the initial experiment. The specific analysis parameters used for each experiment and any changes or additions to the workflow are reported in the following sections.

Titration of aTc concentration in drinking water

Germ-free C57BL/6(J) mice were maintained on the chow diet as described above and received water containing 1, 10, or 30 μg/ml of aTc as well as 100 μg/ml of kanamycin sulfate 1 day prior to gavage of 1×10^9 cells of *E. coli* BL21(DE3) transformed with pFS_0453. Plasmid DNA was extracted and concentrated as described above and 6.25 μl of plasmid DNA was used as an input into SENECA using annealed adapter ligation oligonucleotides FS_0963 and FS_0964 (table S25).

Record-seq comparison of transient chow, fat, and starch-based dietary stimulus

Germ-free C57BL/6(J) mice received the standard chow diet (Kliba Nafag 3307) prior to the experiment. At the beginning of the experiment (2 days before gavage), one group remained on the chow diet, whereas the other two groups received either a starch-based purified diet (Research Diets D12450Jii) or a fat-based (lard-based) purified diet (Research Diets D12492ii), both of which were sterilized by two rounds of irradiation with each 10-20 kGy. Mice received drinking water containing 30 μg/ml of aTc and 100 μg/ml of kanamycin sulfate 1 day before gavage with 1×10^9 cells of *E. coli* MG1655 transformed with pFS_0453 as described above. After 7 days on different diets, mice from all groups received the chow diet (effectively switching fat- and starch-fed mice to the chow diet). Plasmid DNA from fecal pellets was extracted and concentrated as outlined above. Additionally, intestinal contents were sampled yielding the data presented in fig. S1. The initial, 20-day diet switch experiment described in Fig. 1 and fig. S2 used 3.75 μl of plasmid DNA as an input into SENECA using annealed adapter ligation oligonucleotides FS_2759 and FS_2769 (table S25). The consecutive, 14-day diet switch experiment described in fig. S3 used 7.5 μl of plasmid DNA as an input into SENECA using annealed adapter ligation oligonucleotides FS_2240 and FS_2248 (table S25).

Primary analysis was performed for Record-seq and RNA-seq readout as described above for both experiments. During secondary analysis, the minimum value of cumulative transcript-aligning spacer counts (*C*) was set to 10,000 for Record-seq samples, and the minimum value of transcript-aligning counts for RNA-seq samples was set to 100,000. PCA projections and UMAP embeddings were generated for the top 500 most variable genes across days and diets. For the initial 20-day experiment, hierarchical clustering and heatmap generation were performed for DEGs detected by multiple testing on day 7 since this was the last day when the mice were fed different diets prior to switching all mice to the chow diet. For both experiments, diet-specific signature genes were defined as the top 500 DEGs detected for day 7.

Hierarchical clustering and heatmap generation were performed for the final day in each experiment using these diet-specific signature genes. For the initial experiment, genes enriched or depleted in each diet pair comparison were identified on day 7 using pairwise DE testing and used for generating volcano plots. EcoCyc pathway (table S5) enrichment was performed using DEGs identified for each diet pair (tables S2 and S3) and the top hits were used for creating pathway enrichment plots. STRING network analysis was performed using Cytoscape using DEG with $\log_2 FC \geq 1.0$. The granularity parameter for MCL clustering of STRING networks was set at 2.5 and stringdb:: score was used as array source with an edge cutoff of 0.5. Clusters containing more than four nodes were reported. KEGG- and GO-based OA were performed for DEGs identified for each diet pair using clusterProfiler (table S6). Regression analysis was performed to examine the similarity of regulation in the latter experiment for DEGs in mice fed the chow diet or purified diet based on starch detected in the initial experiment.

E. coli Δ*idnK*/Δ*gntK in vivo competition assay*

One group of germ-free C57BL/6(J) mice was switched to the starch-based diet 48 hours before gavage whereas the other remained on the standard chow diet. *E. coli* MG1655 (WT, NCBI Assembly database PRJNA807125) was grown in 200 ml of LB and *E. coli* MG1655 Str^R Δ*idnK*/Δ*gntK* in 250 ml of LB with 30 μg/ml of chloramphenicol (Sigma-Aldrich) at 37°C, 180 rpm overnight. Cultures were pelleted by centrifugation at 3480*g* for 10 min at room temperature, washed twice with 250 ml of sterile PBS and combined after the first washing step at a 1:1 ratio. The combined *E. coli* strains were finally resuspended in 7.5 ml of sterile PBS and diluted 1:10 in sterile PBS to gavage ~1×10^9 CFU per mouse in 500 μl. Feces were collected at 24 hours, 48 hours, and 72 hours after gavage and lysed in 1 ml of PBS using a Retsch MM400 tissue lyser at 30 Hz for 3 min. Serial dilutions were streaked onto LB agar and LB agar containing 30 μg/ml of chloramphenicol and the competitive indices were calculated as $CFU_{\Delta idnK/\Delta gntK}/CFU_{wt} = CFU_{\Delta idnK/\Delta gntK}/(CFU_{total}\text{-}CFU_{\Delta idnK/\Delta gntK})$.

E. coli Δ*uxaC in vivo competition assay*

The competition experiment using the Δ*uxaC* mutant was performed as above with the following alterations: *E. coli* MG1655 Str^R Δ*uxaC* (NCBI Assembly database PRJNA807125) was gavaged into germ-free mice together with either MG1655 WT (NCBI Assembly database PRJNA807125) or with MG1655 Str^R (NCBI Assembly database PRJNA807125). Each strain was grown in 100 ml of LB and the Δ*uxaC* mutant with an additional 50 μg/ml of kanamycin sulfate. Strains were washed with 100 ml of PBS, mixed, and then resuspended in a final volume of 33 ml to achieve a gavage dose of ~1×10^9 CFU per mouse in 500 μl. Serial dilutions of feces were streaked onto LB agar with and without 50 μg/ml of kanamycin sulfate.

Record-seq and RNA-seq assessment of different anatomical sections of the murine gut

One group of germ-free C57BL/6(J) mice was switched to the starch-based diet 48 hours before gavage whereas the other remained on the standard chow diet. Mice received drinking water containing 30 μg/ml of aTc and 100 μg/ml of kanamycin sulfate water and were gavaged with 1×10^9 cells of *E. coli* MG1655 transformed with pFS_0453 as described above. Record-seq plasmid and/or fecal RNA was extracted from individual mouse feces collected daily. On day 7 after gavage, mice were sacrificed and *E. coli* RNA was collected from various intestinal sections of individual mice. Cecal contents were mixed in a Petri dish to homogenize them and then ~100 mg was collected for RNA extraction. Proximal colon contents were collected at a 1-2-cm distance from the cecum. Distal colon contents were collected in the terminal 2 cm of the colon. RNA was extracted, sequenced, and analyzed as described above. The experiment was performed twice: once with samples being pooled from each three individual mice, once with sampling from individual mice. Primary analysis was performed for RNA-seq readout as described above for both experiments. During secondary analysis, the minimum value of transcript-aligning counts was set to 100,000. Differential expression analysis between the diet groups was performed for each intestinal section (cecum, proximal colon and distal colon) independently. Record-seq DEGs were overlapped with DEGs identified through RNA-seq proximally (cecum or proximal colon) or distally (distal colon) ($P_{adj(section)} < 0.1$; $\log_2 FC_{Record-seq} \times \log_2 FC_{section} > 0$). Differential expression analysis between the intestinal sections was performed for each diet group independently. Genes enriched in a particular intestinal section were defined as an overlap of genes identified as up-regulated in that section compared to the other two sections (e.g., genes enriched in the cecum on the chow diet were defined as genes that were up-regulated in the cecum on the chow diet compared to both the proximal and distal colon). Rank-based normalization was used for comparing Record-seq and RNA-seq counts to account for the differences in count distributions.

In vitro exposure of E. coli to DSS

For each replicate, three colonies of *E. coli* MG1655 transformed with pFS_0453 were inoculated into 2 ml of terrific broth (TB) (24 g per liter of yeast extract, 20 g per liter of tryptone, 4 ml per liter of glycerol, 17 mM KH_2PO_4, and 72 mM K_2HPO_4) containing 50 ng/ml of aTc and 0.1, 0.3, 1, 3, or 10% (w/v) of DSS. After overnight culture at 37°C, 220 rpm bacterial cultures were pelleted and plasmid DNA extracted using 500 μl of buffer P1, 500 μl of buffer P2, and 700 μl of buffer N3. Washes were performed as described above. A single elution was performed by adding 60 μl of buffer TE, followed by incubation at 55°C, 850 rpm for 1 min and centrifugation at 20,000*g* for 1 min at room temperature. Plasmid DNA was quantified as described (*18*),

and input normalized for SENECA using annealed adapter ligation oligonucleotides FS_2108 and FS_2109 (table S25). Primary and secondary analysis of data was performed as described above, with the minimum value of cumulative transcript-aligning spacer counts per sample (*C*) set to 10,000.

Record-seq assessment of DSS colitis in vivo

Germ-free C57BL/6(J) mice received drinking water containing 30 μg/ml of aTc and 100 μg/ml of kanamycin sulfate water and were gavaged with 1×10^9 cells of *E. coli* MG1655 or *E. coli* BL21(DE3) transformed with pFS_0453 as described above. DSS (molecular weight 36 to 50 kDa, Lucerna-Chem) was dissolved in tap water to 1, 2, or 3% (w/v) with aTc and kanamycin sulfate as above. Mice received DSS-containing drinking water for 5 days as illustrated in the experimental outlines. Mice were subsequently switched back to aTc and kanamycin sulfate-containing water without DSS. Mice were monitored daily for signs of distress. Mice treated with 3% DSS (w/v) in the drinking water had to be removed from the study on day 13 of the experiment due to colitis-induced distress and are thus omitted from the analysis for days 14 to 19 in Fig. 4, B and C. Plasmid DNA from fecal pellets was extracted and concentrated as outlined above.

The 2% DSS (w/v) experiment (Fig. 4D) with *E. coli* MG1655 used 3.75 μl of plasmid DNA as an input to SENECA using annealed adapter ligation oligonucleotides FS_2806 and FS_2807 (table S25). The initial experiment, including different concentrations of DSS (1%, 2%, 3% w/v) with *E. coli* BL21(DE3) used 7.5 μl of plasmid DNA as an input to SENECA using the annealed adapter ligation oligonucleotides FS_2238 and FS_2246 (table S25).

Primary analysis of the Record-seq sequencing readout for both experiments was performed as described above. For the initial experiment (Fig. 4A) using *E. coli* BL21(DE3), the minimum value of *C* was set to 5,000 counts. Hierarchical clustering and heatmap generation using DEGs was performed for day 19. For inferring the trajectory of DSS-induced differences between treatment groups using PCA-projected data from days 5 to 9 (fig. S6D), an approach analogous to Slingshot (*56*) was employed: clusters were defined using k-medoids clustering with an optimal k value ($k = 5$) determined by the Elbow method based on minimizing within-cluster sum of squares. Branching of trajectories was inferred based on the assumption that clusters with samples from earlier timepoints branch out into clusters with samples from later timepoints, with the nodes of the plotted trajectory indicating the center of each cluster. For discriminating between treatment groups, Record-seq data on days 6 to 19 (treatment and post-treatment) was randomly split into a 70% training and a 30% test set using a fixed random seed to ensure reproducibility. One-versus-rest classification was performed on the test sets with SVM models trained and tuned on the training sets using leave-one-out cross-validation implemented in the e1071 package in R. Receiver operating characteristic (ROC) curves were generated using the ROCR package in R.

For secondary analysis, in the 2% DSS experiment (*E. coli* MG1655), the threshold for *C* was set to 10,000 counts for Record-seq. PCA dimensionality reduction and k-medoids clustering were performed for Record-seq samples collected post DSS treatment. k-medoids cluster identity was encoded by convex hulls in PCA projections. UMAP and time-course $\log_2 FC$ plots was generated and differential expression was performed using days 2 to 20, and identified DEGs were used for EcoCyc pathway enrichment (table S9) and STRING network analysis using Cytoscape. MCL clustering was performed with the granularity parameter set at 2.5 and stringdb::score was used as array source with an edge cutoff of 0.5 to identify groups in the STRING networks. Clusters containing more than two genes were reported. KEGG- and GO-based OA were performed for DEGs using clusterProfiler (table S10).

Record-seq in the presence or absence of B. theta

Germ-free C57BL/6(J) mice received the standard chow diet throughout the entire experiment, drinking water containing 30 μg/ml of aTc and 100 μg/ml of kanamycin sulfate from the day before gavage with 1×10^9 cells of *E. coli* MG1655 transformed with pFS_0453 as described above. For oral gavage of *B. theta*, a single colony of *B. theta* was grown overnight in 140 ml of brain heart infusion broth (Thermo Fisher Scientific) supplemented with 0.5 mg per liter of menadione and 5 mg per liter of hemin (both Sigma-Aldrich) at 37°C under anaerobic conditions without agitation. After two wash steps with 150 ml of sterile anaerobic PBS, *B. theta* was resuspended and mixed with *E. coli* suspension to gavage each 1×10^9 CFU *E. coli* and 1×10^9 CFU *B. theta* in a total volume of 500 μl. Plasmid DNA was extracted from fecal samples using the procedure outlined above but increasing the buffer volumes to 1000 μl of buffer P1, 1000 μl of buffer P2, and 1400 μl of buffer N3. 2-propanol precipitation and resuspension were performed as described above. The 9-day experiment used 3.75 μl of plasmid DNA as an input into the SENECA using annealed adapter ligation oligonucleotides FS_2762 and FS_2772 (table S25). The 27-day experiment used 3.75 μl of plasmid DNA as an input and annealed adapter ligation oligonucleotides FS_2758 and FS_2768 (table S25).

For measurement of bacterial colonization levels, feces were collected, weighed, and lysed in 1 ml of PBS using a Retsch MM400 tissue lyser at 30 Hz for 3 min. Serial dilutions were streaked onto LB agar plates to determine *E. coli* CFU and on BHI agar with 5% defibrinated sheep blood to determine *B. theta* CFU and incubated aerobically or anaerobically, respectively.

Primary analysis of data was performed for Record-seq data as described above for both experiments. For secondary analysis, the minimum value of *C* was set to 5000. This lower threshold for cumulative transcript-aligning spacer counts was chosen due to fewer observed counts, possibly explained by the difficulty in plasmid DNA extraction from *E. coli* colonized with *B. theta*. Hierarchical clustering was performed and heatmaps were generated for the full time-course using DEGs detected on at least 2 days. Combined lists of DEGs identified on any day (table S11 and S12) were used for subsequent EcoCyc pathway enrichment (table S13), STRING network analysis using Cytoscape, and KEGG- and GO-based OA using clusterProfiler (table S14). The granularity parameter for MCL clustering of STRING networks was set at 2.5 and stringdb::score was used as array source with an edge cutoff of 0.5. Clusters containing more than three genes were reported.

Record-seq function in complex microbiomes (sDMDMm2 mice)

C57BL/6(J) mice with the sDMDMm2 microbiota (table S23) received the standard chow diet or the starch-based purified diet (Research Diets D12450Jii) and were switched to drinking water containing 30 μg/ml of aTc but no kanamycin sulfate 6 days before gavage. The gavage procedure was modified as follows: 400 ml of an overnight culture of *E. coli* MG1655 cells transformed with pFS_0453 in LB with 50 μg/ml of kanamycin sulfate were diluted 1:5 into 2 liters pre-warmed LB with 30 ng/μl of aTc and 50 μg/ml of kanamycin sulfate and cultured for another 2 hours. Bacteria were pelleted by centrifugation at 3480*g* for 10 min at room temperature, washed twice with 1 liter of PBS and resuspended in 12 ml to gavage 6×10^{10} CFU into each mouse in 500 μl of PBS. Plasmid DNA was extracted from fecal samples using the procedure outlined above but increasing the buffer volumes to 1000 μl of buffer P1, 1000 μl of buffer P2, and 1400 μl of buffer N3. Isopropanol precipitation and resuspension were performed as described above. SENECA adapter ligation was performed using 3.75 μl of plasmid DNA and annealed adapter ligation oligonucleotides FS_2760 and FS_2770 as well as FS_2761 and FS_2771 (table S25) with the following modification compared to the standard protocol: annealed adapter ligation oligonucleotides were diluted 1:10 instead of 1:100 in TE buffer after annealing. *E. coli* colonization levels were measured by homogenizing

and serial dilution of fecal pellets as described for the *B. theta* experiment above but serial dilutions were spread on MacConkey agar (Thermo Fisher Scientific) with 50 μg/ml of kanamycin sulfate to ensure the selective growth of *E. coli* sentinel cells.

Primary analysis of data was performed for Record-seq data as described above for both experiments. For secondary analysis, the minimum value of *C* was set to 5000. Since the genome-aligning spacer counts increased over the sampling time course, data from the final 21-hour timepoint was used for identifying DEGs, hierarchical clustering and heatmap generation, EcoCyc pathway enrichment (table S17), STRING network analysis using Cytoscape, and KEGG- and GO-based OA using clusterProfiler (table S18). The granularity parameter for MCL clustering of STRING networks was set at 2.5 and stringdb::score was used as array source with an edge cutoff of 0.5. Clusters containing at least two genes were reported.

Transcription-stimulated spacer acquisition

Transcription-stimulated recording constructs were generated by designing gBlock FS_3046 designed following the supplementary methods of (*38*). gBlock FS_3046 (table S25) encodes "T7UUCG(terminator)_rrnBT1(terminator)_T7Term(terminator)_Golden-Gate(BbsI)_FsLeader-CRISPR-02-array-DR2_50bp(stuffer)_EK120029600(terminator_inverted)" and is inserted downstream of *Fs*RT-Cas1–Cas2 in pFS_0453 cloning with XhoI and NotI. This yields a plasmid where oligonucleotides encoding various promoter sequences can be inserted using a Golden Gate reaction with BbsI. Based on pFS_1061, plasmids pFS_1061 to pFS_1067 as well as pFS_1112 to pFS_1114 were generated. These plasmids encode constitutive *E. coli* promoters of different transcriptional activity upstream of the FsLeader_CRISPR_02_array. These promoters are inserted into pFS_1061 by generating double stranded (dsDNA) fragments encoding the respective promoter and appropriate overhangs through the annealing of oligonucleotides FS_3047 and FS_3048 (for pFS_1062), FS_3049 and FS_3050 (for pFS_1063), FS_3051 and FS_3052 (for pFS_1064), FS_3053 and FS_3054 (for pFS_1065), FS_3055 and FS_3056 (for pFS_1066), FS_3057 and FS_3058 (for pFS_1067), FS_3210 and FS_3211 (for pFS_1112), FS_3212 and FS_3213 (for pFS_1113), FS_3214 and FS_3215 (for pFS_1114). Oligonucleotide sequences are listed in table S25. Oligonucleotides were annealed by mixing 2.5 μl of 100 μl oligonucleotides in buffer TE with 5 μl of NEBuffer 2.0 and 40 μl of Ultrapure H_2O (ThermoFisher Scientific) per reaction.

The reaction was heated to 95°C for 5 min in a thermocycler and cooled to 22°C at a rate of 0.5°C/sec. Then the annealed oligonucleotides were diluted 1:200 in Ultrapure H_2O. For each target plasmid a Golden Gate reaction was performed containing 40 fmol of pFS_1061, 1 μl of 1:200 diluted, annealed oligonucleotides, 1 μl of a mixture of ATP and DTT (10 mM each) (ThermoFisher Scientific), 0.25 μl of T7 DNA Ligase (NEB), 0.75 μl of 40% (w/v) PEG8000 (Sigma-Aldrich), 0.75 μl of BpiI (ThermoFisher Scientific), and 1 μl of buffer green (ThermoFisher Scientific). Each reaction was brought up to a 10-μl final volume using Ultrapure H_2O (ThermoFisher Scientific).

From each Golden Gate reaction, 0.5 μl was transformed into 5 μl of chemically competent *E. coli* Stbl3. Individual clones were grown in LB media containing 50 μg/ml of kanamycin sulfate, isolated by plasmid mini-prep and validated by Sanger sequencing (GATC Eurofins). Correct clones were then transformed into chemically competent *E. coli* MG1655 to assess efficiency of recording. In vitro recording and SENECA reaction was performed as described previously (*17*, *18*). Upon identification of J23103 [iGEM Registry of Standard Biological Parts (http://parts.igem.org)] gBlock FS_3344 encoding "T7UUCG(terminator)_rrnBT1(terminator)_T7Term(terminator)_J23103_FsLeader-CRISPR-01-array-DR1_50bp(stuffer)_EK120029600(terminator_inverted)" was designed and cloned into pFS_0453 using XhoI and NotI, yielding pFS_1142.

Record-seq for multiplexed recording in different bacterial chassis

For in vivo multiplexed recording experiments, pFS_1113 and pFS_1142 were transformed into chemically competent MG1655 Str^RΔ*uxaC* ("Δ*uxaC*", NCBI Assembly database PRJNA807125) or MG1655 Str^R Nal^R ("wt", NCBI Assembly database PRJNA807125) and spread on LB-Agar plates containing 50 μg/ml of kanamycin sulfate. Germ-free C57BL/6(J) mice were switched to the starch-based diet 72 hours before gavage and received drinking water containing 30 μg/ml of aTc and 100 μg/ml of kanamycin sulfate from 24 hours before gavage onwards. All *E. coli* strains were inoculated into LB medium with 100 μg/ml of streptomycin sulfate and 50 μg/ml of kanamycin sulfate with or without 50 μg/ml of nalidixic acid, respectively and grown overnight at 37°C with 180 rpm shaking. Following washing as described above, the following strains were mixed 1:1 and a total of 1×10^{10} CFU (5×10^9 CFU per strain and mouse) were orally gavaged into recipient mice. Figure 7B, wt-pFS_1113 together with Δ*uxaC* pFS_1142 and wt-pFS_1142 together with Δ*uxaC* pFS_1113. Colonization levels of the strains were determined by fecal dilution and plating on LB/Str100/Kan50 (both Δ*uxaC* and wt) and LB/Str100/Kan50/Nal50 (only wt).

Plasmid DNA was extracted from fecal samples using the procedure outlined above but increasing the buffer volumes to 500 μl of buffer P1, 500 μl of buffer P2, and 700 μl of buffer N3. 2-propanol precipitation and resuspension were performed as described above.

SENECA adapter ligation was performed using 7.5 μl of plasmid DNA and annealed adapter ligation with annealed oligonucleotides FS_3194 and 3204 as well as FS_3316 and 3321 (table S25) with the following modification compared to the standard protocol: annealed adapter ligation oligonucleotides were diluted 1:10 instead of 1:100 in TE buffer after annealing. Upon annealing to dsDNA fragments, oligonucleotides FS_3194 and 3204 form an overhang compatible with FaqI-digested *Fs*DR2 and oligonucleotides FS_3316 and 3321 form an overhang compatible with FaqI-digested *Fs*DR1. Therefore, both sets of annealed oligos were used in a single SENECA adapter ligation reaction to simultaneously read out CRISPR spacer acquisition into pFS_1113 and pFS_1142. Accordingly, SENECA first-round PCR was performed with FS_968, FS_969, FS_970, FS_971, FS_972, FS_973, FS_974 described previousl, which are primers compatible with FsDR2 along with additional primers compatible with FsDR1, namely FS_3325, FS_3326, FS_3327, FS_3328, FS_3329, FS_3330, FS_3331 (table S25) and were each used at a concentration of 0.714 μM. The universal reverse primer FS_911 (*18*) was used at a concentration of 10 μM. SENECA second-round PCR was performed as described above.

For primary analysis of data, a modified version of the Snakemake workflow described above was used. This workflow requires the additional input of a table with sample-specific entries for the reference plasmid, DR sequence and library barcode. Samples with multiple DR-barcoded *E. coli* strains, such as those from the in vivo multiplexed recording experiment, resulted in reads containing both DR sequences (*Fs*DR2 and *Fs*DR1). These were stratified into strain-specific reads by identifying the DR-specific library barcode using fuzzy string matching and processed independently. Secondary analysis was performed as described above.

For the in vitro multiplexed recording experiment, the minimum value of cumulative transcript-aligning spacer counts (*C*) was set to 30,000. For the in vivo multiplexed recording experiment, the minimum value of cumulative transcript-aligning spacer counts (*C*) was set to 10,000. DEGs were identified between the strains (wt vs Δ*uxaC*) for days 7 to 10 in both groups of mice independently (group 1: Δ*uxaC*-DR1, wt-DR2; group 2: wt-DR1, Δ*uxaC*-DR2; n = 5 in each group). High-confidence DEGs were defined as DEGs identified as regulated in the same direction for at least 3 out of 4 days in both comparisons. These high-confidence DEGs (table S19) were used for EcoCyc pathway enrichment (table S20) and KEGG- and GO-based OA using clusterProfiler (table S21).

REFERENCES AND NOTES

1. R. U. Sheth, H. H. Wang, DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* **19**, 718–732 (2018). doi: 10.1038/s41576-018-0052-8; pmid: 30237447
2. F. Farzadfard, T. K. Lu, Emerging applications for DNA writers and molecular recorders. *Science* **361**, 870–875 (2018). doi: 10.1126/science.aat9249; pmid: 30166483
3. F. Schmidt, R. J. Platt, Applications of CRISPR-Cas for synthetic biology and genetic recording. *Curr. Opin. Syst. Biol.* **5**, 9–15 (2017). doi: 10.1016/j.coisb.2017.05.008
4. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016). doi: 10.1126/science.aaf1175; pmid: 27284167
5. S. Ishiguro, H. Mori, N. Yachie, DNA event recorders send past information of cells to the time of observation. *Curr. Opin. Chem. Biol.* **52**, 54–62 (2019). doi: 10.1016/j.cbpa.2019.05.009; pmid: 31200335
6. L. Yang *et al*., Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014). doi: 10.1038/nmeth.3147; pmid: 25344638
7. C. Munck, R. U. Sheth, D. E. Freedberg, H. H. Wang, Recording mobile DNA in the gut microbiota using an Escherichia coli CRISPR-Cas spacer acquisition platform. *Nat. Commun.* **11**, 95 (2020). doi: 10.1038/s41467-019-14012-5; pmid: 31911609
8. D. T. Riglar, P. A. Silver, Engineering bacteria for diagnostic and therapeutic applications. *Nat. Rev. Microbiol.* **16**, 214–225 (2018). doi: 10.1038/nrmicro.2017.172; pmid: 29398705
9. M. R. Charbonneau, V. M. Isabella, N. Li, C. B. Kurtz, Developing a new class of engineered live bacterial therapeutics to treat human diseases. *Nat. Commun.* **11**, 1738 (2020). doi: 10.1038/s41467-020-15508-1; pmid: 32269218
10. D. T. Riglar *et al*., Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. *Nat. Biotechnol.* **35**, 653–658 (2017). doi: 10.1038/nbt.3879; pmid: 28553941
11. E. J. Archer, A. B. Robinson, G. M. Süel, Engineered *E. coli* that detect and respond to gut inflammation through nitric oxide sensing. *ACS Synth. Biol.* **1**, 451–457 (2012). doi: 10.1021/sb3000595; pmid: 23656184
12. J. W. Kotula *et al*., Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4838–4843 (2014). doi: 10.1073/pnas.1321321111; pmid: 24639514
13. M. Mimee, A. C. Tucker, C. A. Voigt, T. K. Lu, Programming a human commensal bacterium, bacteroides thetaiotaomicron, to sense and respond to stimuli in the murine gut microbiota. *Cell Syst.* **2**, 214 (2016). doi: 10.1016/j.cels.2016.03.007; pmid: 27135367
14. G. P. Donaldson, S. M. Lee, S. K. Mazmanian, Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2016). doi: 10.1038/nrmicro3552; pmid: 26499895
15. N. Zmora *et al*., Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features. *Cell* **174**, 1388–1405.e21 (2018). doi: 10.1016/j.cell.2018.08.041; pmid: 30193112
16. Q. Tang *et al*., Current sampling methods for gut microbiota: A call for more precise devices. *Front. Cell. Infect. Microbiol.* **10**, 151 (2020). doi: 10.3389/fcimb.2020.00151; pmid: 32328469
17. F. Schmidt, M. Y. Cherepkova, R. J. Platt, Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380–385 (2018). doi: 10.1038/s41586-018-0569-1; pmid: 30283135
18. T. Tanna, F. Schmidt, M. Y. Cherepkova, M. Okoniewski, R. J. Platt, Recording transcriptional histories using Record-seq. *Nat. Protoc.* **15**, 513–539 (2020). doi: 10.1038/s41596-019-0253-4; pmid: 31925399
19. I. M. Keseler *et al*., The EcoCyc database: Reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Res.* **45** (D1), D543–D550 (2017). doi: 10.1093/nar/gkw1003; pmid: 27899573
20. L. V. Hooper, J. I. Gordon, Glycans as legislators of host-microbial interactions: Spanning the spectrum from symbiosis to pathogenicity. *Glycobiology* **11**, 1R–10R (2001). doi: 10.1093/glycob/11.2.1R; pmid: 11287395
21. E. C. Martens, H. C. Chiang, J. I. Gordon, Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008). doi: 10.1016/j.chom.2008.09.007; pmid: 18996345
22. N. J. Sweeney, D. C. Laux, P. S. Cohen, Escherichia coli F-18 and E. coli K-12 eda mutants do not colonize the streptomycin-treated mouse large intestine. *Infect. Immun.* **64**, 3504–3511 (1996). doi: 10.1128/iai.64.9.3504-3511.1996; pmid: 8751891
23. M. P. Leatham *et al*., Mouse intestine selects nonmotile flhDC mutants of Escherichia coli MG1655 with increased colonizing ability and better utilization of carbon sources. *Infect. Immun.* **73**, 8039–8049 (2005). doi: 10.1128/IAI.73.12.8039-8049.2005; pmid: 16299298
24. T. Conway, P. S. Cohen, Commensal and pathogenic *Escherichia coli* metabolism in the gut. *Microbiol. Spectr.* **3**, 3.3.24 (2015). doi: 10.1128/microbiolspec.MBP-0006-2014; pmid: 26185077
25. M. Ricci, Laboratory animal control diets: Very important, often neglected. *Lab Anim. (NY)* **44**, 240–241 (2015). doi: 10.1038/laban.786
26. Y. Kim, T. K. Wood, Toxins Hha and CspD and small RNA regulator Hfq are involved in persister cell formation through MqsR in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **391**, 209–213 (2010). doi: 10.1016/j.bbrc.2009.11.033; pmid: 19909729
27. M. Kitagawa, Y. Matsumura, T. Tsuchido, Small heat shock proteins, IbpA and IbpB, are involved in resistances to heat and superoxide stresses in *Escherichia coli*. *FEMS Microbiol. Lett.* **184**, 165–171 (2000). doi: 10.1111/j.1574-6968.2000.tb09009.x; pmid: 10713416
28. L. G. Patwa *et al*., Chronic intestinal inflammation induces stress-response genes in commensal Escherichia coli. *Gastroenterology* **141**, 1842–51.e1, 10 (2011). doi: 10.1053/j.gastro.2011.06.064; pmid: 21726510
29. M. E. Johansson *et al*., Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model. *PLOS ONE* **5**, e12238 (2010). doi: 10.1371/journal.pone.0012238; pmid: 20805871
30. S. A. Cevallos *et al*., Increased epithelial oxygenation links colitis to an expansion of tumorigenic bacteria. *mBio* **10**, e02244-19 (2019). doi: 10.1128/mBio.02244-19; pmid: 31575772
31. J. J. Lemke *et al*., Direct regulation of *Escherichia coli* ribosomal protein promoters by the transcription factors ppGpp and DksA. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5712–5717 (2011). doi: 10.1073/pnas.1019383108; pmid: 21402902
32. K. M. Ng *et al*., Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–99 (2013). doi: 10.1038/nature12503; pmid: 23995682
33. M. S. Desai *et al*., A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339–1353.e21 (2016). doi: 10.1016/j.cell.2016.10.043; pmid: 27863247
34. A. Marcobal *et al*., Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **10**, 507–514 (2011). doi: 10.1016/j.chom.2011.10.007; pmid: 22036470
35. B. Yilmaz *et al*., Long-term evolution and short-term adaptation of microbiota strains and sub-strains in mice. *Cell Host Microbe* **29**, 650–663.e9 (2021). doi: 10.1016/j.chom.2021.02.001; pmid: 33662276
36. S. Brugiroux *et al*., Genome-guided design of a defined mouse microbiota that confers colonization resistance against *Salmonella enterica* serovar Typhimurium. *Nat. Microbiol.* **2**, 16215 (2016). doi: 10.1038/nmicrobiol.2016.215; pmid: 27869789
37. M. Shakya, C. C. Lo, P. S. G. Chain, Advances and challenges in metatranscriptomic analysis. *Front. Genet.* **10**, 904 (2019). doi: 10.3389/fgene.2019.00904; pmid: 31608125
38. J. B. Budhathoki *et al*., Real-time observation of CRISPR spacer acquisition by Cas1-Cas2 integrase. *Nat. Struct. Mol. Biol.* **27**, 489–499 (2020). doi: 10.1038/s41594-020-0415-7; pmid: 32367067
39. G. Nemoz, J. Robert-Baudouy, F. Stoeber, Physiological and genetic regulation of the aldohexuronate transport system in *Escherichia coli*. *J. Bacteriol.* **127**, 706–718 (1976). doi: 10.1128/jb.127.2.706-718.1976; pmid: 783117
40. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6640–6645 (2000). doi: 10.1073/pnas.120163297; pmid: 10829079
41. Y. Uchimura *et al*., Complete genome sequences of 12 species of stable defined moderately diverse mouse microbiota 2. *Genome Announc.* **4**, e00951-16 (2016). doi: 10.1128/genomeA.00951-16; pmid: 27634994
42. H. Li *et al*., The outer mucus layer hosts a distinct intestinal microbial niche. *Nat. Commun.* **6**, 8292 (2015). doi: 10.1038/ncomms9292; pmid: 26392213
43. M. B. Stead *et al*., RNAsnap™: A rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Res.* **40**, e156 (2012). doi: 10.1093/nar/gks680; pmid: 22821568
44. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). doi: 10.1093/bioinformatics/btu170; pmid: 24695404
45. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: 10.1038/nmeth.1923; pmid: 22388286
46. H. Li *et al*., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi: 10.1093/bioinformatics/btp352; pmid: 19505943
47. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014). doi: 10.1093/bioinformatics/btt656; pmid: 24227677
48. J. Köster, S. Rahmann, Snakemake—A scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012). doi: 10.1093/bioinformatics/bts480; pmid: 22908215
49. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
50. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for dimension reduction (2018).
51. N. T. Doncheva, J. H. Morris, J. Gorodkin, L. J. Jensen, Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* **18**, 623–632 (2019). doi: 10.1021/acs.jproteome.8b00702; pmid: 30450911
52. J. H. Morris *et al*., clusterMaker: A multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* **12**, 436 (2011). doi: 10.1186/1471-2105-12-436; pmid: 22070249
53. Gene Ontology Consortium, The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res.* **49** (D1), D325–D334 (2021). doi: 10.1093/nar/gkaa1113; pmid: 33290552
54. M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, M. Tanabe, KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49** (D1), D545–D551 (2021). doi: 10.1093/nar/gkaa970; pmid: 33125081
55. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012). doi: 10.1089/omi.2011.0118; pmid: 22455463
56. K. Street *et al*., Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018). doi: 10.1186/s12864-018-4772-0; pmid: 29914354

ACKNOWLEDGMENTS

We are grateful to C. Beisel, E. Burcklen, I. Nissen, K. Eschbach, M. Feldkamp, and T. Schär from the Genomics Facility Basel and T. Leeb and P. Nicholson of the Next-Generation Sequencing Facility of the University of Bern for assistance in Illumina sequencing, as well as H.-M. Kaltenbach, G. Rätsch, and the entire Platt and Macpherson laboratories for discussing results. S. Misztal provided experimental support, and J. Limenitakis undertook *E. coli* whole-genome sequencing analysis. **Funding:** This work was supported by the European Research Council grant 851021 (F.S., T.T, and R.J.P.); European Union Horizon 2020 Marie Skłodowska-Curie grant 744257 (J.Z.); European Research Council grant 742195 (J.Z and A.J.M.); ETH Zurich grant to Gunnar Rätsch 1-001785-000 (T.T.); Botnar Research Centre for Child Health Multi-Investigator Project Grant (R.F., A.J.M., and R.J.P.); Public Health Service grant GM117324 (T.C.); Swiss National Science Foundation grants CRSII5_177164, CRSII3_154414, and 310030_179479 (A.J.M.); National Centres of Competence – Molecular Systems Engineering 51NF40-182895 (A.J.M. and R.J.P.); Personalized Health and Related Technologies grant 2021-542 (A.J.M. and R.J.P.); and Brain & Behavior Research Foundation grant 26606, Personalized Health and Related Technologies grant PHRT-203, Swiss National Science Foundation grant 31003A_175830, ETH Zurich grant ETH-27 18-2, Swiss Cancer Research Foundation grant KFS-4863-08-2019, Botnar Research Centre for Child Health Fast Track Call grant, EMBO grant 4217, and J. Fickel (R.J.P.). The Clean Mouse Facility is supported by the Genaxen Foundation, Inselspital, and the University of Bern. **Author contributions:** F.S., J.Z., T.T., A.J.M., and R.J.P. conceived of the study and planned experiments; F.S. and J.Z. performed experiments; T.T., F.S., and J.Z. curated data; T.T. wrote software packages; T.T., F.S., and J.Z. visualized data; F.S., J.Z., T.T., A.J.M., and R.J.P. analyzed data and wrote the manuscript; R.F. provided advice, comments and contributed to responses to the reviewers; T.C. contributed to planning the methodology of the competitive

colonization experiments and provided the mutant *E. coli* strains; and A.J.M. and R.J.P. supervised the study and acquired funding.
Competing interests: F.S., J.Z., T.T., A.J.M., and R.J.P. are co-inventors on patent applications filed by ETH Zurich related to this work. All other authors have no competing interests. **Data and materials availability:** Plasmid pFS_0453 (#117006), pFS_1061 (#184738), pFS_1064 (#184739), pFS_1112 (#184740), pFS_1113 (#184741) and pFS_1142 (#184742) used for recording experiments in this study are available from Addgene. Deep sequencing data of RNA-seq and Record-seq experiments is available in the National Center for Biotechnology Information Sequence Read Archive (PRJNA807096). Whole-genome sequencing of *E. coli* strains used in this study is available in the National Center for Biotechnology Information Assembly database (PRJNA807125). The datasets generated and/or analyzed during the current study are available from the corresponding authors upon reasonable request. Software tools, custom scripts, and R notebooks describing workflows are available on the Platt lab GitHub (https://github.com/plattlab).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm6038
Figs. S1 to 12
Tables S1 to 25

Submitted 1 October 2021; accepted 15 March 2022
10.1126/science.abm6038

RESEARCH ARTICLE SUMMARY

NEUROIMMUNOLOGY

Reversible CD8 T cell–neuron cross-talk causes aging-dependent neuronal regenerative decline

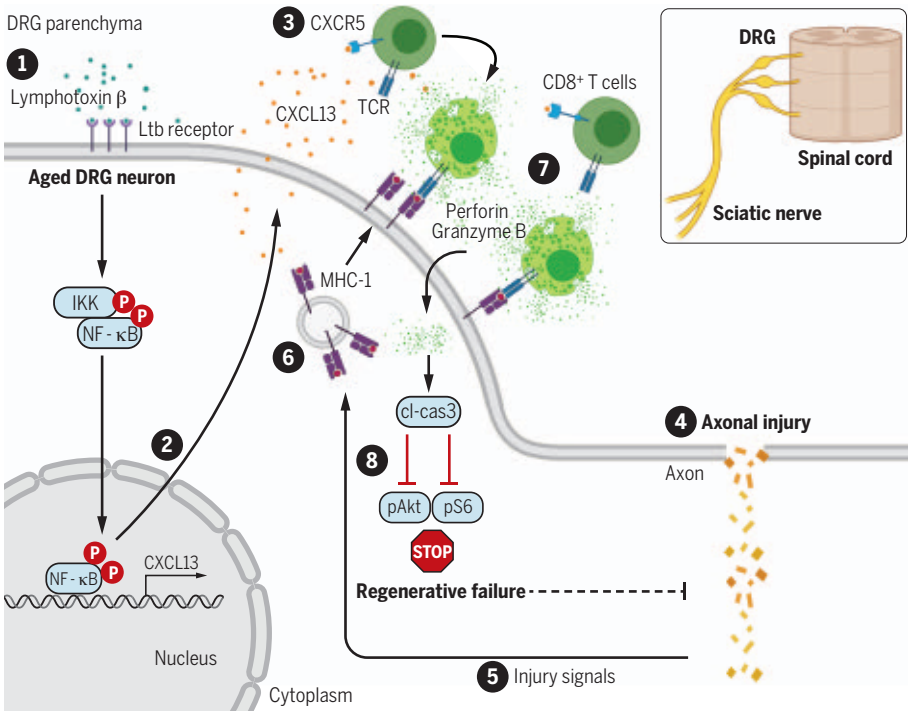
Luming Zhou†, Guiping Kong†, Ilaria Palmisano, Maria Teresa Cencioni, Matt Danzi, Francesco De Virgiliis, Jessica S. Chadwick, Greg Crawford, Zicheng Yu, Fred De Winter, Vance Lemmon, John Bixby, Radhika Puttagunta, Joost Verhaagen, Constandina Pospori, Cristina Lo Celso, Jessica Strid, Marina Botto, Simone Di Giovanni*

INTRODUCTION: Axonal regeneration and neurological functional recovery are extremely limited in the elderly. Consequently, injuries to the nervous system are typically followed by severe and long-term disability. Our understanding of the molecular mechanisms underlying aging-dependent regenerative failure is poor, hindering progress in the development of effective therapies for neurological repair. To facilitate the design of repair strategies, there is a pressing need to identify critical molecular and cellular mechanisms that cause regenerative failure in aging.

RATIONALE: Aging causes a broad spectrum of modifications in cell signaling, including changes in metabolism, immunity, and overall tissue homeostasis, which play key roles in nervous system physiology and response to insults. Thus, we hypothesized that injuries to the aged nervous system would be followed by unique molecular and cellular modifications that would contribute to aging-dependent regenerative decline. To this end, molecular and cellular signatures associated with aging and injury to the nervous system were systematically investigated by performing RNA sequencing from dorsal root ganglia (DRG) in a well-established model of sciatic nerve injury in young versus aged mice. Insight into these mechanisms could allow the discovery of previously unrecognized molecular targets to counteract aging-dependent regenerative decline.

RESULTS: Initial analysis of RNA sequencing data identified that aging was mainly associated with a marked increase in T cell activation and signaling in DRG after sciatic nerve injury in mice. Subsequent experiments demonstrated that aging was associated with increased inflammatory cytokines including lymphotoxins in DRG both preceding and following sciatic nerve injury. Specifically, we found that lymphotoxin β was required for the phosphorylation of NF-κB that drives the expression of the chemokine CXCL13 in DRG sensory neurons. CXCL13 attracted $CD8^+$ T cells that expressed the CXCL13 receptor CXCR5 in proximity to neurons that act as antigen-presenting cells by overexpressing major histocompatibility complex class I (MHC I) after a sciatic nerve injury. The engagement of $CXCR5^+CD8^+$ T cells with MHC I–expressing sensory neurons activated caspase 3, which impaired pAKT and pS6 signaling, leading to regenerative failure. Pharmacological antagonism of caspase 3 activation reversed regenerative failure in aged animals and restored pAKT and pS6 expression. Adoptive transfer experiments including comparison between *Cxcr5*$^{-/-}$ and wild-type cells directly implicated $CXCR5^+CD8^+$ T cells in restricting axonal regeneration after sciatic injury in the aged animals. Finally, neutralization of CXCL13 with monoclonal antibodies antagonized the recruitment of $CXCR5^+CD8^+$ T cells to the DRG and restored the regenerative ability of sciatic sensory axons in aged mice to levels comparable to those found in young animals. CXCL13 antagonism also significantly promoted skin reinnervation and neurological recovery of sensory function.

CONCLUSION: We have identified an aging-dependent mechanism that relies on the CXCL13-dependent recruitment and activation of $CXCR5^+CD8^+$ T cells in proximity to sensory neurons that restrict axonal regeneration after nerve injury after communication with MHC I–presenting DRG neurons. CXCL13 antagonism with humanized monoclonal antibodies can revert regenerative decline and promote neurological recovery in aged mice, suggesting a pathway that could be exploited in future therapies. ■



Sequence of events leading to aging-dependent regenerative failure in injured sensory neurons. (1) Lymphotoxin β activates NF-κB; (2) phosphorylated NF-κB increases CXCL13 expression; (3) $CXCR5^+CD8^+$ T cells are recruited into the DRG parenchyma; (4 and 5) sciatic axonal injury activates injury signals; (6) MHC I is presented by DRG neurons; (7) $CXCR5^+CD8^+$ T cells engage with MHC I and activate neuronal caspase 3; (8) cleaved caspase 3 reduces pAKT and pS6, leading to regenerative failure.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: s.di-giovanni@imperial.ac.uk
†These authors contributed equally to this work.
Cite this article as L. Zhou *et al.*, *Science* **376**, eabd5926 (2022). DOI: 10.1126/science.abd5926

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd5926

RESEARCH ARTICLE

NEUROIMMUNOLOGY

Reversible CD8 T cell–neuron cross-talk causes aging-dependent neuronal regenerative decline

Luming Zhou[1]†, Guiping Kong[1]†, Ilaria Palmisano[1], Maria Teresa Cencioni[2], Matt Danzi[3], Francesco De Virgiliis[1], Jessica S. Chadwick[1], Greg Crawford[4], Zicheng Yu[1], Fred De Winter[5], Vance Lemmon[3], John Bixby[3], Radhika Puttagunta[6], Joost Verhaagen[5], Constandina Pospori[7,8], Cristina Lo Celso[7,8], Jessica Strid[4], Marina Botto[4], Simone Di Giovanni[1]*

Aging is associated with increased prevalence of axonal injuries characterized by poor regeneration and disability. However, the underlying mechanisms remain unclear. In our experiments, RNA sequencing of sciatic dorsal root ganglia (DRG) revealed significant aging-dependent enrichment in T cell signaling both before and after sciatic nerve injury (SNI) in mice. Lymphotoxin activated the transcription factor NF-κB, which induced expression of the chemokine CXCL13 by neurons. This in turn recruited $CXCR5^+CD8^+$ T cells to injured DRG neurons overexpressing major histocompatibility complex class I. $CD8^+$ T cells repressed the axonal regeneration of DRG neurons via caspase 3 activation. CXCL13 neutralization prevented $CXCR5^+CD8^+$ T cell recruitment to the DRG and reversed aging-dependent regenerative decline, thereby promoting neurological recovery after SNI. Thus, axonal regeneration can be facilitated by antagonizing cross-talk between immune cells and neurons.

With aging, there is a decline in axonal regenerative ability that undermines recovery and often leads to long-term disability (*1*–*4*). The cellular and molecular mechanisms underlying aging-dependent regenerative decline remain poorly understood. Previous studies have shown that an age-related impairment in dedifferentiation and activation of Schwann cells limits axonal regrowth in the injured peripheral nervous system, impairing sensory and motor recovery (*5*, *6*).

Aging itself produces a wide range of profound modifications in cell signaling, metabolism, immunity, gene regulation, and protein translation, which affect tissue homeostasis (*7*–*9*). Therefore, we hypothesized that aging involves the development of unique molecular and cellular mechanisms preceding injury, which, in combination with injury-specific signals, would prime neurons toward regenerative failure. A better understanding of these mechanisms could allow the identification of additional molecular strategies to counteract aging-dependent regenerative decline.

[1]Division of Neuroscience, Department of Brain Sciences, Imperial College London, London, UK. [2]Division of Neurology, Department of Brain Sciences, Imperial College London, London, UK. [3]Miami Project to Cure Paralysis, Department of Neurological Surgery, Miller School of Medicine, University of Miami, Miami, FL, USA. [4]Department of Immunology and Inflammation, Imperial College London, London, UK. [5]Netherlands Institute for Neuroscience, Royal Academy of Arts and Sciences, Amsterdam, Netherlands. [6]Spinal Cord Injury Center, Heidelberg University Hospital, Heidelberg, Germany. [7]Haematopoietic Stem Cell Laboratory, Francis Crick Institute, London, UK. [8]Department of Life Sciences, Imperial College London, South Kensington Campus, London, UK.
*Corresponding author. Email: s.di-giovanni@imperial.ac.uk
†These authors contributed equally to this work.

RNA sequencing (RNA-seq) of sciatic nerve dorsal root ganglia (DRG) in mice identified an aging-dependent enrichment in immune and cytokine/chemokine signaling that intensified after sciatic nerve injury. The aging-dependent increase in inflammatory cytokines, including lymphotoxin, activated NF-κB in DRG neurons, which up-regulated expression of the chemokine CXCL13. CXCL13 in turn recruited $CXCR5^+CD8^+$ T cells to sites near neurons acting as antigen-presenting cells (APCs) by overexpressing major histocompatibility complex class I (MHC I). Activated $CD8^+$ T cells repressed axonal regeneration of sensory DRG neurons by inhibiting regenerative signals via caspase 3–dependent reduction in pAKT and pS6. In vivo $CD8^+$ T cell depletion or CXCL13 neutralization restored axonal regeneration of sensory neurons promoting functional recovery.

Thus, we identify an aging-dependent mechanism restricting the regenerative ability of axons. Moreover, we propose antibody-mediated manipulation of neuron–immune cell cross-talk as a clinically suitable approach for neurological repair in the elderly.

Aging induces extensive changes in adaptive immune gene profiles in sciatic DRG at homeostasis and after injury

RNA-seq was performed on sciatic DRG from 8- to 10-week-old (young) versus 20- to 22-month-old (aged) mice either preceding (sham) or following a sciatic nerve injury (SNI). Changes in gene expression were compared to the baseline expression in young sham animals. Significantly differentially expressed genes [false discovery rate (FDR) < 0.05] were analyzed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) for functional classification (Fig. 1A, fig. S1A, and data S1 and S2). The highest number of significantly differentially expressed genes (FDR < 0.05) was identified in aged DRG after injury (fig. S1, B to D, and data S1). The down-regulation of GO classes implicated in neurodevelopment, neurotransmitters, ion transport, G protein–coupled signaling, and signal transduction observed in young animals failed to occur in aged DRG even after SNI (Fig. 1A and data S2). Aged DRG displayed a marked enrichment in gene expression associated with adaptive immune responses, especially T cell and cytokine/chemokine signaling, which was amplified after SNI (Fig. 1A, fig. S1A, and data S2). The prominent aging-dependent regulation of the immune response was then prioritized for further analysis. A shortlist of 310 transcripts (data S3) enriched for immune response according to GO and KEGG ($P < 0.01$) was used to map hypothetical signaling networks. The network revealed interconnected functional clusters including chemokine-cytokine interactions and B and T cell signaling (Fig. 1B). Notably, the aging-dependent T cell responses coincided with a major increase in the number of T cell receptor (TCR) clonotypes (fig. S1E).

Neuronal CXCL13 is elevated in aged DRG along with $CD8^+$ T cells expressing CXCR5

Cxcl13 was the most prominently up-regulated gene after sciatic injury (Fig. 1C and data S3). The expression of the CXCL13 receptor *Cxcr5* was also significantly enhanced in aged DRG (Fig. 1C and data S3), suggesting the presence of a CXCL13-CXCR5 signaling axis. A significant increase in CXCL13 expression was also observed in aged sciatic DRG neurons both at baseline and 3 days after SNI, when regeneration is reduced relative to young DRG neurons (Fig. 1, D to H, and fig. S1, F to H). Although Cxcl13 expression was also detected around the sciatic nerve injury site, no aging-dependent changes were observed (fig. S1, I and J). CXCL13 protein in DRG 3 days after SNI also significantly increased in the aged DRG group relative to young DRG mice (Fig. 1I).

Because CXCL13 is a chemoattractant for $CXCR5^+$ B and T cells, we next assessed the localization and immunophenotype of B and T cells within the DRG 3 days after SNI in young or aged mice. B and T cells including $CXCR5^+$ B cells, $CXCR5^+CD8^+$ T cells, and $CD4^+$ T cells significantly increased in aged DRG parenchyma (Fig. 2, A to D, and fig. S2, A to F).

A considerable percentage of $CXCR5^+CD8^+$ T cells were $CD44^+CD69^+CD62L^-$, indicating effector memory T cells (Fig. 2, E and F, and fig. S2D), whereas only a very small proportion of $CD8^+$ T cells or $CXCR5^+CD8^+$ T cells were $CD103^+$ and $P2RX7^+$, suggesting a migratory rather than a primary tissue-resident phenotype. Approximately half of both $CD8^+$ and

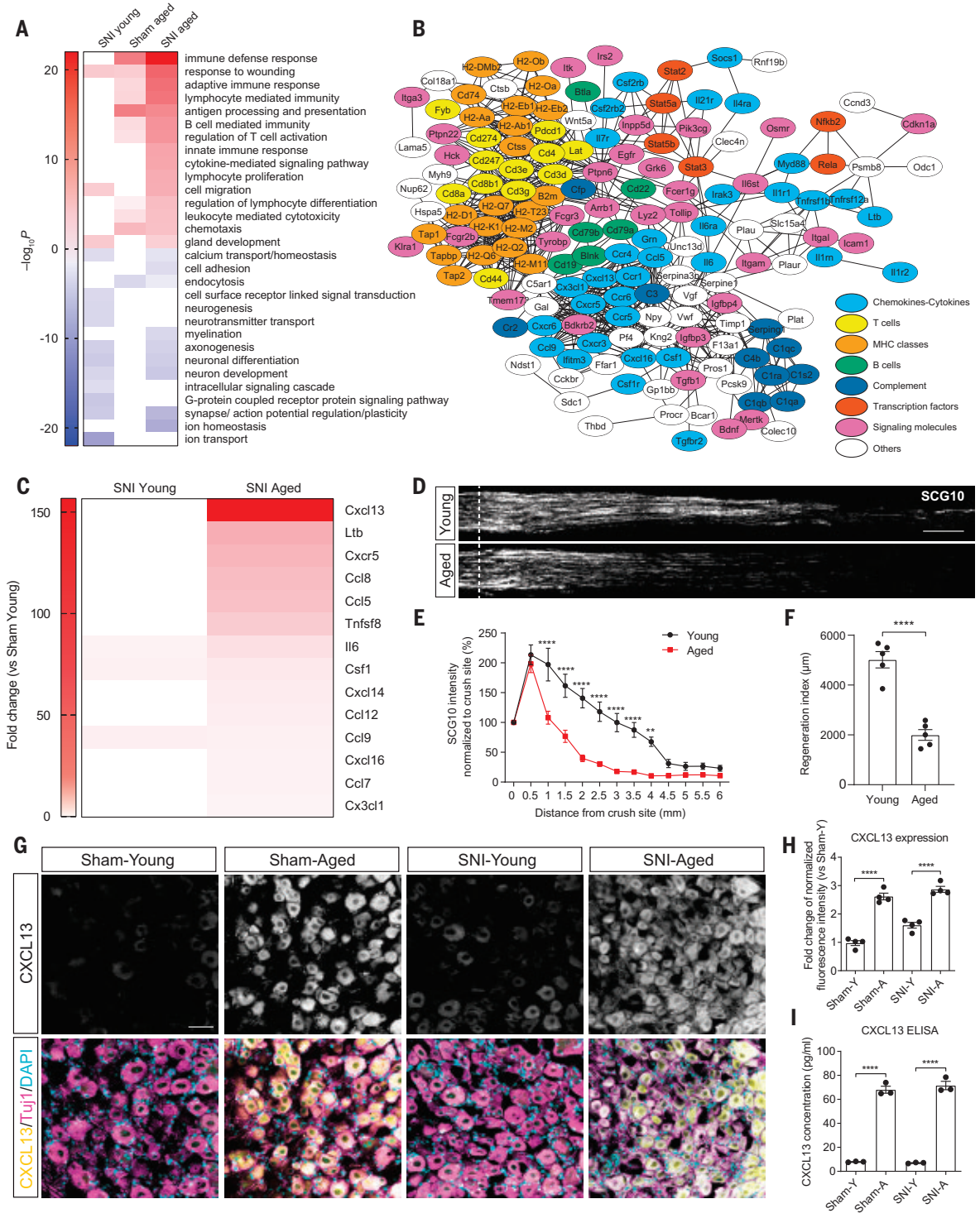


Fig. 1. Aging induces enrichment in genes associated with chemokine and cytokine expression and adaptive immunity in dorsal root ganglia, including after SNI. (**A**) Heatmap shows the top 15 categories of Gene Ontology Biological Processes (GO-BPs) of the differentially expressed genes (FDR < 0.05) from three comparisons: SNI young, sham aged, and SNI aged versus sham young, respectively (n = 3). GO-BPs enrichment of up-regulated (red) and down-regulated (blue) are indicated by $-\log_{10}P$ (P < 0.01). (**B**) Cytoscape visualization of the STRING protein-protein interaction network of the up-regulated genes (FDR < 0.05) involved in immune response in SNI aged versus sham young (P < 0.01). Each node is a gene, color-coded according to STRING functional annotation. (**C**) Heatmap shows the fold change of the differentially expressed chemokines and cytokines. (**D**) Representative SCG10 immunostaining of sciatic nerves 3 days after SNI. The dashed line indicates the proximal crush site. Scale bar, 500 μm. (**E**) Normalized SCG10 intensity to the proximal crush site (n = 5). (**F**) Analysis of the regeneration index (n = 5). (**G**) Representative immunostaining of CXCL13, the neuronal marker Tuj1, and DAPI in DRG 3 days after SNI. Scale bar, 50 μm. (**H**) Quantification of CXCL13 expression shown in (G) (n = 4). (**I**) CXCL13 ELISA from sciatic DRG (n = 3). **P < 0.01, ****P < 0.0001 [two-way ANOVA with post hoc Sidak's test in (E); unpaired Student's t test in (F); one-way ANOVA with post hoc Tukey's test in (H) and (I)]. Data are means ± SEM and represent three [(A), (C), and (I)] or two [(E), (F), and (H)] experiments.

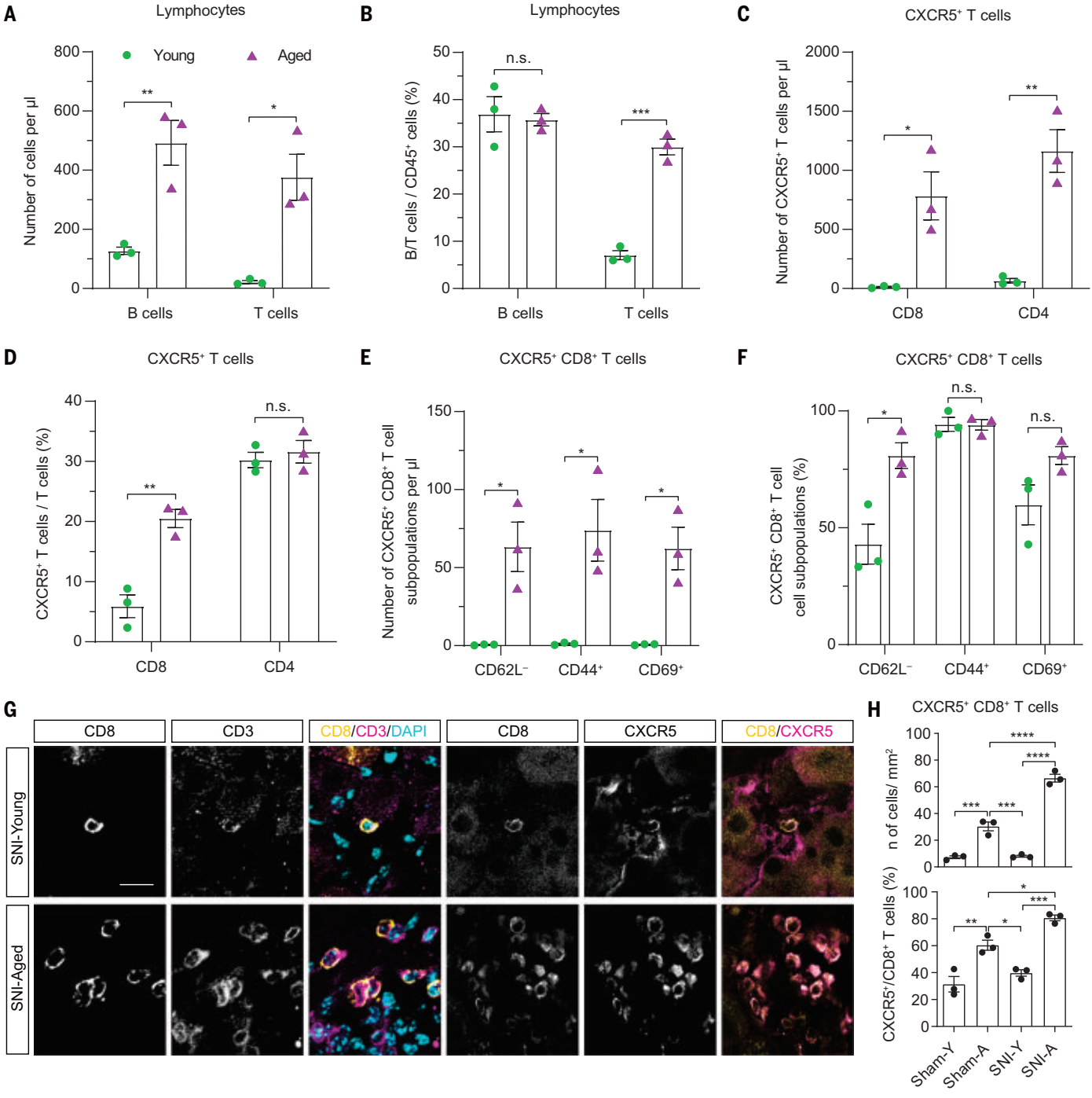


Fig. 2. CXCR5$^+$CD8$^+$ T cell characterization in DRG before and after SNI. (**A**) Quantification of B and T cell number normalized to the counting beads in DRG (n = 3). (**B**) Percentage of B and T cells of total CD45$^+$ cells in DRG (n = 3). (**C**) Number of CXCR5$^+$CD8$^+$ or CD4$^+$ T cells normalized to the counting beads in DRG (n = 3). (**D**) Percentage of CXCR5$^+$CD8$^+$ or CD4$^+$ T cells in DRG (n = 3). (**E**) Number of CD62L$^-$, CD44$^+$, or CD69$^+$/CXCR5$^+$CD8$^+$ T cells normalized to the counting beads in DRG (n = 3). (**F**) Percentage of CD62L$^-$, CD44$^+$, or CD69$^+$/CXCR5$^+$CD8$^+$ T cells in DRG (n = 3). (**G**) CD8 coimmunostained with CD3 and DAPI or CXCR5 in DRG 3 days after SNI. Scale bar, 10 µm. (**H**) Quantification of CXCR5$^+$CD8$^+$ T cell number and percentage of CXCR5$^+$CD8$^+$ T cells of total CD8$^+$ T cells in DRG in sham or 3 days after SNI (n = 3). *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 [unpaired Student's t test in (A) to (F) or one-way ANOVA with post hoc Tukey's test in (H)]; n.s., not significant. Data are means ± SEM and represent three experiments.

CXCR5$^+$CD8$^+$ T cells also expressed CXCR6, suggesting cell retention and restricted recirculation ability (fig. S3).

In aged animals, the percentage of naïve total T cells in the blood was reduced, whereas the percentage of memory CXCR5$^+$ T cells was significantly enhanced (fig. S4). The number of CD8$^+$ T cells, including CXCR5$^+$CD8$^+$ T cells, significantly increased in aged sciatic DRG relative to young at baseline, with a further significant increase 3 days after SNI (Fig. 2, G and H). All CD8$^+$ cells were CD68$^-$ and CD3$^+$, indicating that they were T cells and not macrophages (fig. S5A). Finally, B cell numbers in aged DRG after SNI increased, but these cells

localized outside of the DRG parenchyma, likely in the perivenular space (fig. S5, B and C). Thus, aging is associated with enhanced neuronal expression of CXCL13 and recruitment of CXCR5$^+$ active and resident T cells to the DRG parenchyma before and after SNI.

Neuronal CXCL13 is sufficient to recruit CXCR5$^+$ B and T cells to the DRG and to impair axonal regeneration

Next, we asked whether CXCL13 was secreted and whether its overexpression was sufficient to recruit CXCR5$^+$ B and T cells to the DRG. We first transduced cultured DRG neurons with adeno-associated virus (AAV) to overexpress CXCL13, and culture medium was collected for the subsequent in vitro splenocyte migration assays. Medium from CXCL13-overexpressing cultures versus control enhanced the recruitment of CXCR5$^+$ B and T cells (fig. S6).

To address whether CXCL13 expression would affect axonal regeneration, we overexpressed CXCL13 in vivo in DRG neurons, and interferon-γ (IFN-γ) and mannitol were delivered systemically to induce MHC I presentation and enhance blood-DRG barrier permeability, respectively. Immediately thereafter, a SNI was performed (fig. S7). Both CXCR5$^+$ B and CXCR5$^+$CD8$^+$ T cells were significantly increased in *Cxcl13*-overexpressing DRG relative to green fluorescent protein (GFP) control (Fig. 3, A to F). Sciatic nerve axonal regeneration was significantly reduced after *Cxcl13* overexpression (Fig. 3, G to I).

Thus, CXCL13 expressed by DRG neurons can recruit CXCR5$^+$ B and T cells and can inhibit axonal regeneration, partially phenocopying the aging phenotype.

CXCL13 expression in aged DRG requires lymphotoxin-dependent NF-κB phosphorylation

To identify the transcription factors (TFs) putatively involved in driving aging-dependent CXCL13 expression, we asked whether TFs previously reported to promote CXCL13 expression were induced in aged DRG in our RNA-seq dataset. Among the TFs known to bind to the CXCL13 promoter—including *Hif1a*,

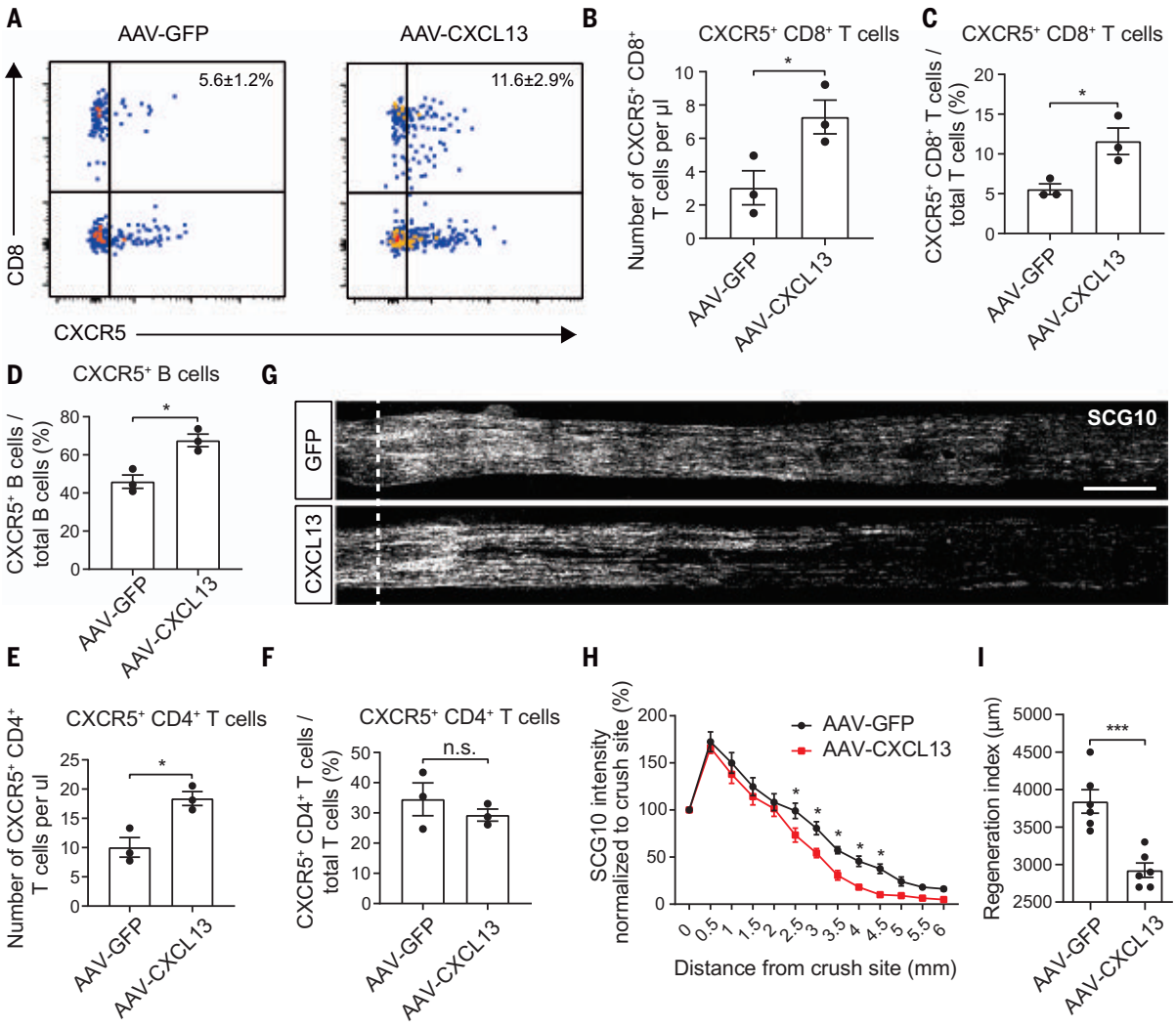


Fig. 3. Overexpression of CXCL13 in DRG neurons in young mice facilitates the recruitment of CXCR5$^+$ B and T cells to the DRG and reduces sciatic nerve regeneration. (**A**) Flow cytometry analysis of CXCR5$^+$CD8$^+$ T cells in AAV-GFP– or AAV-*Cxcl13*–infected sciatic DRG 3 days after SNI. (**B**) Quantification of CXCR5$^+$CD8$^+$ T cell number normalized to the counting beads in the sciatic DRG (n = 3). (**C**) Percentage of CXCR5$^+$CD8$^+$ T cells in sciatic DRG (n = 3). (**D**) Percentage of CXCR5$^+$ B cells in sciatic DRG (n = 3). (**E**) Number of CXCR5$^+$CD4$^+$ T cells in sciatic DRG normalized to the counting beads (n = 3). (**F**) Percentage of CXCR5$^+$CD4$^+$ T cells in sciatic DRG (n = 3). (**G**) Representative SCG10 immunostaining of sciatic nerves from AAV-GFP– or AAV-*Cxcl13*–infected mice. The dashed line indicates the proximal injury site. Scale bar, 500 µm. (**H**) Normalized SCG10 intensity to the proximal crush site from the mice infected with AAV-GFP and AAV-*Cxcl13* (n = 6). (**I**) Analysis of the regeneration index (n = 6). *P < 0.05, ***P < 0.001 [unpaired Student's t test in (B), (C), (D), (E), (F), and (I) or two-way ANOVA with post hoc Sidak's test in (H)]. Data are means ± SEM and represent three experiments.

Pou5f1 (*Oct4*), *Rar*, *Nfkb*, and *Sp1* (*10*, *11*)—only *Nfkb2* was significantly up-regulated in aged DRG, which made it a candidate for further investigation (fig. S8A). We next treated cultured DRG neurons with canonical NF-κB activators [tumor necrosis factor–α (TNF-α) or interleukin-1β (IL-1β)] and noncanonical NF-κB activators (BAFF or LTα1β2) for 24 hours, followed by reverse transcription polymerase chain reaction (RT-PCR) for *Cxcl13*, *Nfkb1*, and *Nfkb2*. TNF-α and IL-1β were able to up-regulate the expression of *Nfkb1* and/or *Nfkb2* but not *Cxcl13*, whereas BAFF affected neither *Nfkb* nor *Cxcl13* gene expression (fig. S8B).

However, LTα1β2, which is known to activate noncanonical NF-κB2 (*12*), significantly enhanced the expression of both *Nfkb2* and *Cxcl13*. This effect could be attenuated via the selective inhibition of the NF-κB kinase IKK using the IKK inhibitor PS1145 (fig. S8C). Both NF-κB2 expression and phosphorylation were induced with aging, but phosphorylation of NF-κB2 was inhibited by PS1145 relative to a vehicle control (fig. S8D). Accordingly, aging-induced expression of *Cxcl13* was significantly decreased by PS1145 (fig. S8E). Finally, immunostaining for pNF-κB2 demonstrated neuronal localization within the DRG and a significant increase in aged DRG neurons both before and 3 days after SNI (fig. S8, F and G). Thus, age-related expression of CXCL13 is driven by NF-κB2.

CD8$^+$ T cells and neuronal MHC I are required for aging-dependent regenerative decline after sciatic nerve injury

Next, we asked whether CD8$^+$ T cells, CD4$^+$ T cells, or B cells are required to impair axonal regeneration of aged sciatic DRG neurons. When CD8$^+$ T cells were depleted using a CD8α monoclonal antibody (mAb) administered to aged mice 1 week before SNI, we observed significantly improved nerve regeneration (Fig. 4, A to C) and a significant reduction of CD8$^+$ T cells in the DRG relative to animals that received an isotype control (fig. S9). By contrast, the depletion of CD4$^+$ T or B cells did not alter aging-dependent regenerative decline after SNI (fig. S10).

These data implicate CD8$^+$ T cells in the regenerative decline that follows SNI in aged animals. RNA-seq revealed an aging-dependent enrichment in the expression of MHC I after SNI (Fig. 1B). MHC I expression increased before and after SNI in aged DRG neurons relative to young controls (Fig. 4, D to F). MHC I expression did not show any selectivity for neuronal subtypes (fig. S11, A to C). However, MHC I was not expressed in axons, but rather was restricted to macrophages in proximity to the sciatic nerve lesion site (fig. S11D).

Next, we asked whether antigen-sensitized CD8$^+$ T cells were required to limit neurite outgrowth in MHC I–expressing DRG neurons in culture. MHC I was induced in DRG neurons but absent in glial cells after IFN-γ injection (fig. S12, A to D). TCR-transgenic OT-I CD8$^+$ T cells effected reduced neurite outgrowth in DRG neurons expressing MHC I in the presence of the ovalbumin (OVA) peptide $OVA_{257–264}$ (fig. S12, E and F) as well as an increase in cleaved caspase 3 without an increase in apoptotic features (fig. S12, G and H). The purity of isolated CD8$^+$ T cells from the spleen of OT-I mice was confirmed by flow cytometry (fig. S12I). Thus, the presence of both neuronal MHC I expression and OT-I antigen-specific CD8$^+$ T cells is needed to reduce neurite outgrowth in DRG neurons.

We next explored whether MHC I–dependent peptide presentation was needed for axonal regenerative decline in aging mice. To this end, we used AAV to conditionally express the virally encoded glycine/alanine-rich peptide sequence GAr, which inhibits MHC I antigen presentation, thereby evading CD8$^+$ T cell immune responses (*13*, *14*), in sciatic DRG of aged mice (fig. S13A). A sciatic nerve crush injury was then performed and animals were killed 3 days later (fig. S13B). Neuronal MHC I and cleaved caspase 3 expression were significantly reduced after AAV-GAr-rtTA infection relative to control AAV-rtTA (Fig. 4, G and H). Sciatic axonal regeneration was significantly enhanced upon AAV-GAr-rtTA infection (Fig. 4, I and J, and fig. S13, C to E). Thus, both CD8$^+$ T cells and neuronal MHC I are required for aging-dependent regenerative decline after SNI.

CD8$^+$ T cell–dependent activation of caspase 3 inhibits pAKT and pS6 and is required for aging-dependent axonal regenerative failure

We aimed to characterize the molecular signaling underpinning CD8$^+$ T cell–dependent regenerative decline in aged animals. Cleaved caspase 3 expression significantly increased in DRG neurons of aged mice 3 days after SNI (Fig. 4, K and L) in the absence of DNA fragmentation associated with apoptosis (fig. S14A). Perforin and granzyme B expression significantly increased in aged DRG neurons after SNI (fig. S14, B to D). Active caspase 3 can lead to AKT truncation and reduced AKT phosphorylation (*15*), which in turn reduces S6 phosphorylation. pAKT and pS6 are associated with axonal regeneration (*16*, *17*) and are typically impaired by aging (*18*). Indeed, phosphorylation of AKT and S6 decreased in aged DRG 3 days after SNI. However, this effect was significantly reversed after depletion of CD8$^+$ T cells (Fig. 4, M and N). CD8$^+$ T cell depletion significantly enhanced pAKT and pS6 in DRG neurons while reducing the expression of cleaved caspase 3, perforin, and granzyme B (fig. S14, E to G, and fig. S15, A and B).

Finally, the inhibition of caspase 3 with Z-DEVD-FMK significantly induced AKT and S6 phosphorylation (fig. S15, C and D) and increased sciatic axon regeneration (fig. S15, E to G). Thus, CD8$^+$ T cell–dependent activation of caspase 3 is required to inhibit pAKT and pS6 in aged DRG neurons after SNI. Moreover, the inhibition of caspase 3 activation promotes axonal regeneration.

CXCR5$^+$CD8$^+$ T cells drive aging-dependent axonal regenerative decline after sciatic nerve injury

Next, we sought to determine whether CXCR5$^+$ CD8$^+$ T cells are directly responsible for axonal regenerative decline after SNI. CXCR5$^+$CD8$^+$ T cells were isolated and adoptively transferred to young or aged OT-I *Rag2*$^{-/-}$ mice, which lack B and T cells (fig. S16A). Prior to cell transfer, we found a significant increase in nerve regeneration from aged OT-I *Rag2*$^{-/-}$ mice with and without CD8 T cell depletion relative to wild-type mice, with no change between the two OT-I *Rag2*$^{-/-}$ groups (fig. S16, B to E). Thus, endogenous OT-I CD8$^+$ T cells had no effect on nerve regeneration.

We then asked whether adoptive transfer of CXCR5$^+$CD8$^+$ T cells was sufficient to restrict sciatic nerve regeneration in aged mice after SNI. Sorted CXCR5$^+$CD8$^+$ T cells (fig. S17, A and B) were injected into young or aged OT-I *Rag2*$^{-/-}$ recipients. Adoptively transferred CXCR5$^+$CD8$^+$ T cells robustly infiltrated aged DRG, but not young recipients (Fig. 5, A and B). By contrast, transferred CXCR5$^+$CD8$^+$ T cells accumulated more abundantly in spleens of young mice versus aged mice (Fig. 5, A and B). These migratory patterns of CXCR5$^+$CD8$^+$ T cells correlated with the differential CXCL13 levels in the DRG and spleens (Fig. 5C). Additionally, transferred CXCR5$^+$CD8$^+$ T cells were increased in the blood of aged mice (fig. S17, C and D). Finally, adoptive transfer of CXCR5$^+$CD8$^+$ T cells led to a significantly reduced sciatic nerve regeneration after injury in aged mice alone (Fig. 5, D and E, and fig. S17E). CXCR5$^+$CD8$^+$ T cell transfer resulted in a significant increase of perforin and a decrease of pS6 expression in aged DRG neurons (Fig. 5, F and G), whereas no changes were noted in the DRG of young mice (fig. S17, F to I).

CXCL13 antagonism blocks the recruitment of CXCR5$^+$CD8$^+$ T cells to rescue aging-dependent axonal regenerative decline after sciatic nerve injury

Having demonstrated that recruitment of CXCR5$^+$CD8$^+$ T cells causes regeneration failure, we asked how these cells are recruited. Sorted wild-type or *Cxcr5*$^{-/-}$ CD8$^+$ T cells were transferred to OT-I *Rag2*$^{-/-}$ recipient mice that were treated with control immunoglobulin G (IgG) or CXCL13 mAb before SNI. *Cxcr5*$^{-/-}$ CD8$^+$ T cell recruitment to the DRG was significantly lower than for wild-type CD8$^+$ T cells. CXCL13 mAb significantly reduced the migration

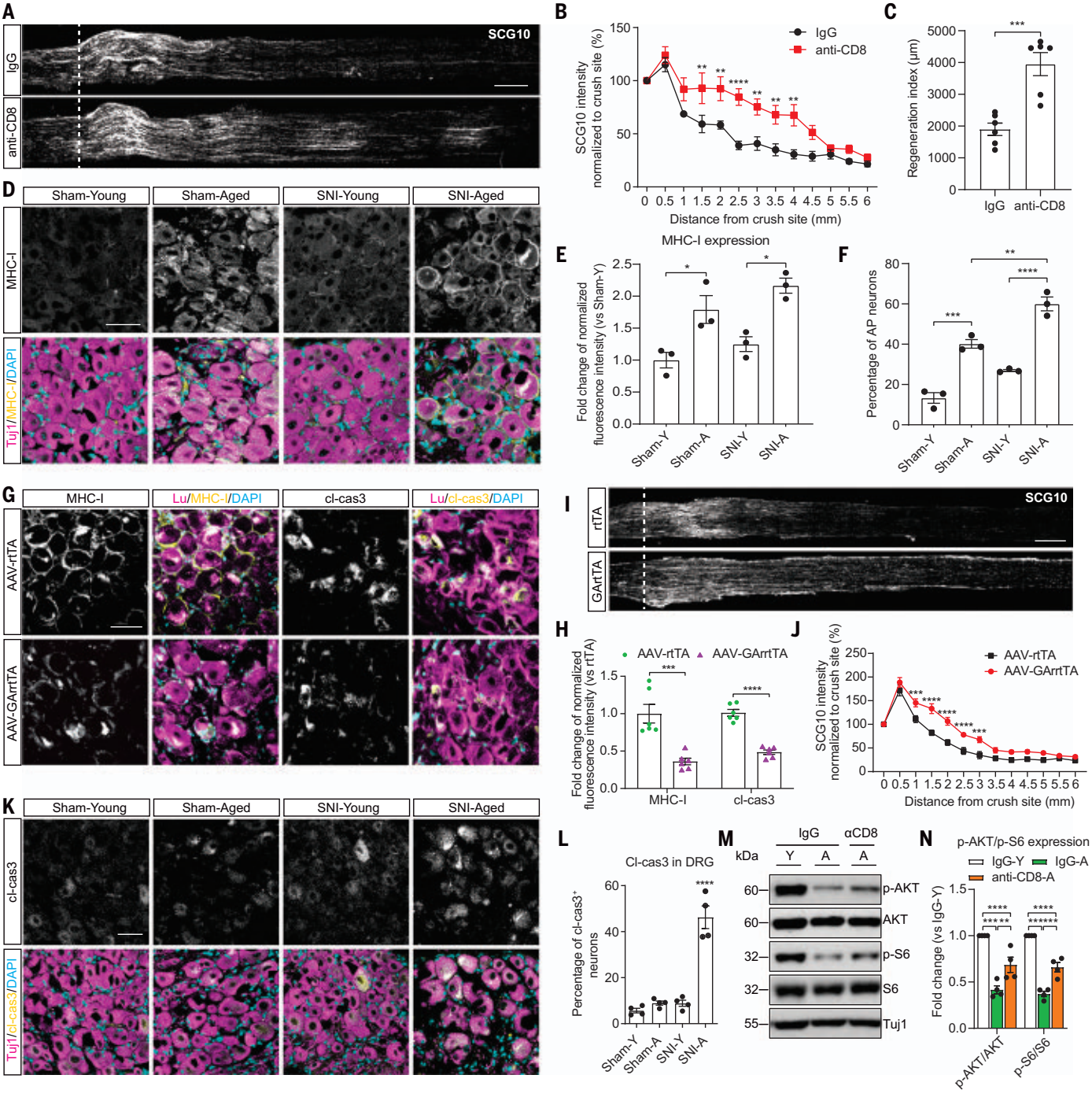


Fig. 4. CD8$^+$ T cells and neuronal MHC I are required for age-dependent regenerative decline after SNI through pAKT and pS6. (**A**) Representative SCG10 immunostaining of sciatic nerves 3 days after SNI after control IgG or CD8 mAb. The dashed line indicates the proximal crush site. Scale bar, 500 μm. (**B**) Normalized SCG10 intensity to the proximal crush site (n = 6). (**C**) Analysis of regeneration index (n = 6). (**D**) Representative MHC I and Tuj1 coimmunostaining and DAPI in sciatic DRG before and 3 days after SNI. Scale bar, 50 μm. (**E**) Fold change of MHC I expression shown in (D) (n = 3). (**F**) Percentage of antigen-presenting (AP) DRG neurons (n = 3). (**G**) Representative MHC I or cleaved caspase 3 (cl-cas3) coimmunostaining with luciferase (Lu) and DAPI in DRG after AAV-rtTA or AAV-GAr-rtTA and 3 days after SNI. Scale bar, 50 μm. (**H**) Fold change of MHC I and cleaved caspase 3 expressions in DRG shown in (G) (n = 6). (**I**) SCG10 immunostaining of sciatic nerves from AAV-rtTA– or AAV-GAr-rtTA– infected aged mice 3 days after SNI. The dashed line indicates the proximal injury site. Scale bar, 500 μm. (**J**) Normalized SCG10 intensity to the proximal crush site (n = 6). (**K**) Representative coimmunostaining of cleaved caspase 3, Tuj1, and DAPI in DRG before and 3 days after SNI. Scale bar, 50 μm. (**L**) Quantification of the percentage of cleaved caspase 3–positive DRG neurons (n = 4). (**M**) Immunoblot shows p-AKT, AKT, p-S6, S6, and Tuj1 expression in the sciatic DRG after control IgG or CD8 mAb 3 days after SNI. (**N**) Immunoblot quantification of p-AKT/AKT and p-S6/S6 in the sciatic DRG (n = 4). *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 [two-way ANOVA with post hoc Sidak's test in (B) and (J); unpaired Student's t test in (C) and (H); one-way ANOVA with post hoc Tukey's test in (E), (F), (L), and (N)]. Data are means ± SEM and represent two [(B), (C), (L), and (N)] or three [(E), (F), (H), and (J)] experiments.

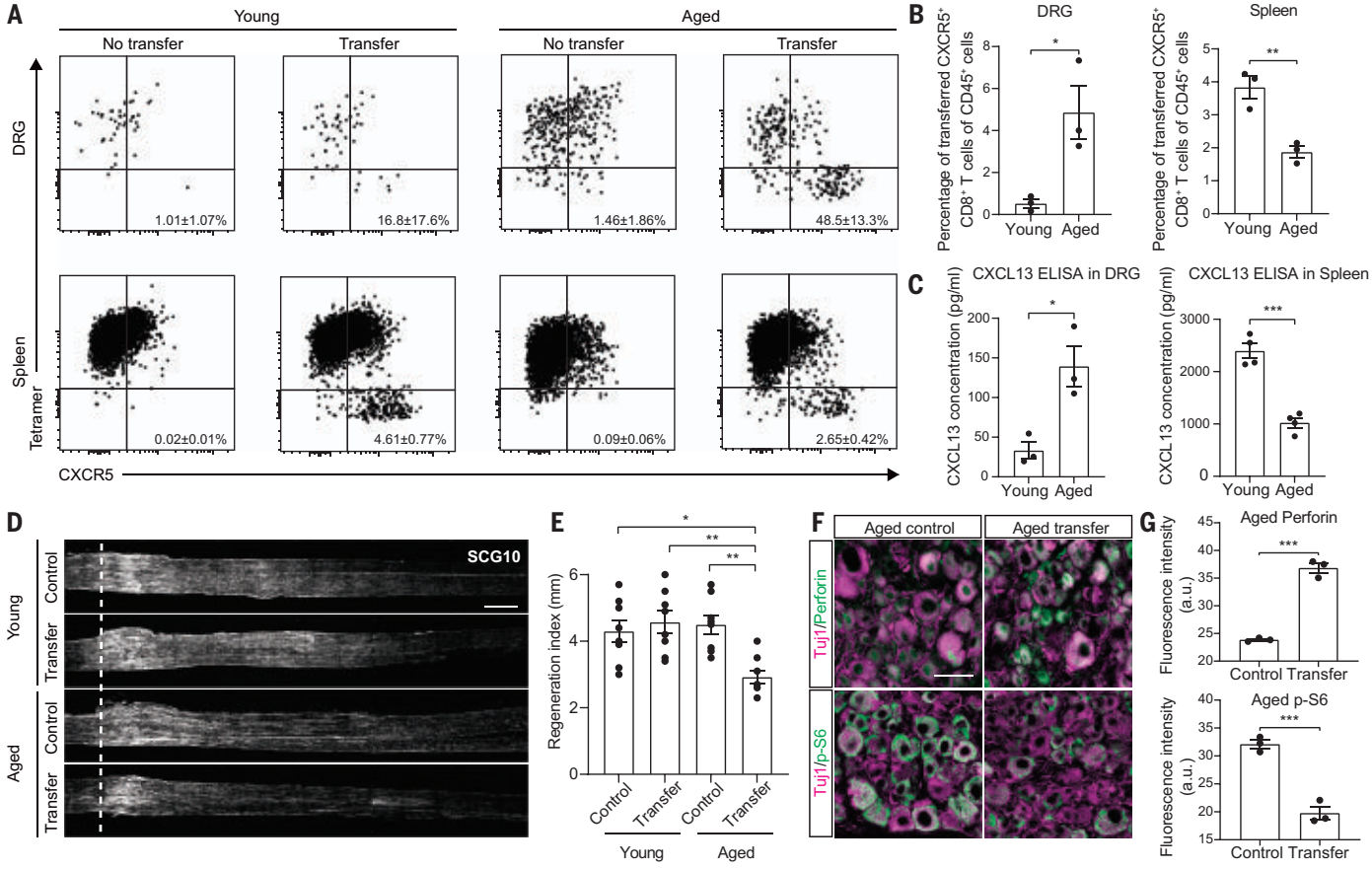


Fig. 5. CXCR5$^+$CD8$^+$ T cells cause aging-dependent regenerative decline after SNI. (**A**) Representative flow cytometry plots of endogenous and transferred CXCR5$^+$CD8$^+$ T cells (gated TCRβ$^+$CD8$^+$ cells) from the sciatic DRG and spleen of OT-I *Rag2*$^{-/-}$ mice 3 days after SNI. Tetramer (SIINFEKL) was used to distinguish the host OT-I CD8 T cells from donated wild-type CD8 T cells. (**B**) Quantification of transferred CXCR5$^+$CD8$^+$ T cells from DRG or spleen 3 days after SNI (n = 3). (**C**) CXCL13 ELISA from OT-I *Rag2*$^{-/-}$ DRG or spleen (n = 3 or 4). (**D**) Representative SCG10 immunostaining of sciatic nerves with or without CXCR5$^+$CD8$^+$ T cell transfer 3 days after SNI in OT-I *Rag2*$^{-/-}$ mice. The dashed line indicates the proximal injury site. Scale bar, 500 μm. (**E**) Analysis of regeneration index (n = 6 or 8). (**F**) Perforin or p-S6 counterstained with Tuj1 in the DRG from aged OT-I *Rag2*$^{-/-}$ mice 3 days after SNI with or without CXCR5$^+$CD8$^+$ T cell transfer. Scale bar, 50 μm. (**G**) Quantification of perforin or p-S6 normalized intensity (n = 3). *P < 0.05, **P < 0.01, ***P < 0.001 [one-way ANOVA with post hoc Tukey's test in (E); unpaired Student's t test in (B), (C), and (G)]. Data are means ± SEM and represent three experiments.

of CXCR5$^+$CD8$^+$ cells to DRG (Fig. 6, A to E) and the spleen (Fig. 6, I and J). Sciatic nerve regeneration was significantly enhanced by CXCL13 antagonism after adoptive transfer of wild-type CD8$^+$ T cells (Fig. 6, F to H), but not of *Cxcr5*$^{-/-}$ CD8$^+$ T cells, which were also associated with higher regeneration relative to wild-type CD8$^+$ T cells (Fig. 6, F to H). Thus, CXCL13 acting on its receptor CXCR5 is required for CD8 T cell migration and to restrict sciatic nerve regeneration in aged DRG.

CXCL13 neutralization reverses aging-dependent regenerative decline and promotes neurological functional recovery after sciatic nerve injury

Finally, we investigated whether CXCL13 neutralization promotes axonal regeneration, epidermal innervation, and functional recovery in aged animals. Injection of CXCL13 mAb significantly enhanced neurite outgrowth in DRG ex vivo culture, reduced the recruitment of CXCR5$^+$ T and B cells, and blocked aging-dependent regenerative failure of sciatic nerve and DRG infiltration of CD8 T cells 3 days after SNI in aged mice but had no effect in young mice (Fig. 7, A to E, and fig. S18, A to F). However, CXCL13 neutralization did not alter the number of CD11b$^+$F4-80$^+$ macrophages in DRG (fig. S19).

In a separate experiment, we tested the effects of CXCL13 antagonism on functional recovery. Aged mice recovered significantly more slowly than young mice, and CXCL13 neutralization induced a significantly accelerated recovery in aged mice in all the sensory modalities investigated (Fig. 7, F to I, and fig. S18, G to I). CXCL13 neutralization promoted significant skin reinnervation from the hairless interdigital area of the hindpaw in aged mice (Fig. 7, J and K, and fig. S18, J and K). Thus, CXCL13 neutralization promotes axonal regeneration, skin reinnervation, and functional recovery in aged animals.

Discussion

Our work indicates that aging primes DRG neurons for regenerative failure after axonal injury by a CXCL13-dependent mechanism that regulates CXCR5$^+$CD8$^+$ T cell entry into the DRG. Regenerative failure is a consequence of the CXCL13-dependent DRG recruitment of CXCR5$^+$CD8$^+$ T cells that communicate with DRG neurons via MHC I on the neuronal cell surface to drive regenerative inhibitory signaling. Unlike in the aged DRG tissue microenvironment, young adult DRG do not harbor T cells before an injury and the T cells do not colonize the DRG parenchyma until many days after nerve injury, past the initiation of the axonal regenerative response (*19*). CXCL13 neutralization reduced the number of CXCR5$^+$CD8$^+$ T cells into the DRG, allowing for significantly

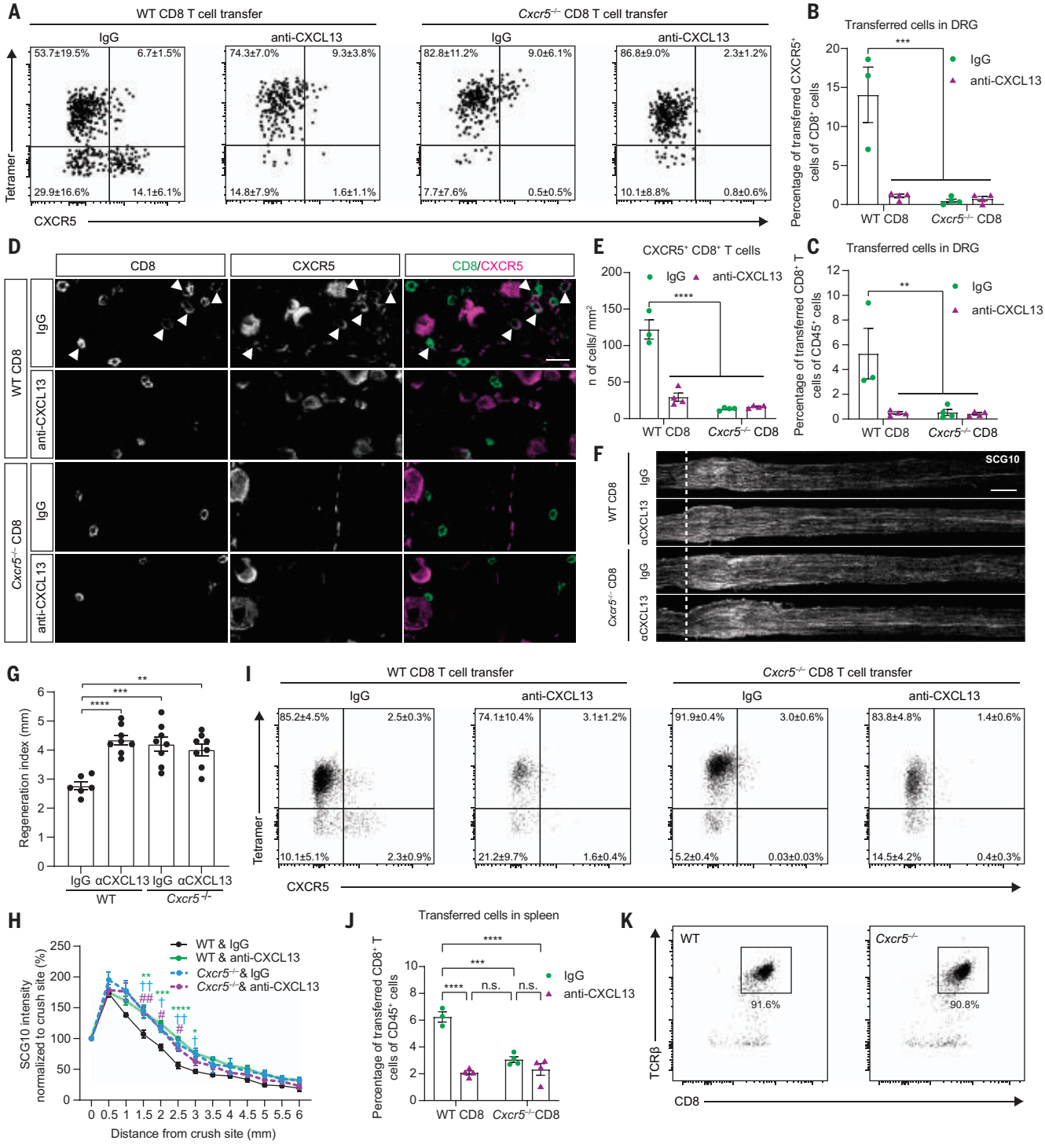


Fig. 6. CXCR5 is required for CD8 T cell infiltration in aged DRG and regenerative decline after SNI. (**A**) Representative flow cytometry plots of endogenous and transferred CXCR5$^+$CD8$^+$ T cells (gated TCRβ$^+$CD8$^+$ cells) from the sciatic DRG of OT-I *Rag2*$^{-/-}$ aged mice 3 days after SNI and treatment with IgG or CXCL13 mAb. (**B**) Quantification of transferred CXCR5$^+$CD8$^+$ T cells from DRG shown in (A) (n = 3 or 4). (**C**) Quantification of transferred CD8$^+$ T cells from DRG (n = 3 or 4). (**D**) Representative coimmunostaining of CD8 and CXCR5 in sciatic DRG 3 days after SNI. Scale bar, 10 μm. (**E**) Quantification of the number of CXCR5$^+$CD8$^+$ T cells in sciatic DRG (n = 4). (**F**) Representative SCG10 immunostaining of sciatic nerves after control IgG or CXCL13 mAb treatment 3 days after SNI. The dashed line indicates the proximal injury site. Scale bar, 500 μm. (**G**) Analysis of regeneration index (n = 6 or 8). (**H**) Normalized SCG10 intensity to the proximal crush site (n = 6 or 8). *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 (wild-type & anti-CXCL13 versus wild-type & IgG); †P < 0.05, ††P < 0.01 (CXCR5$^{-/-}$ & IgG versus wild-type & IgG); #P < 0.05, ##P < 0.01 (CXCR5$^{-/-}$ & anti-CXCL13 versus wild-type & IgG). (**I**) Representative flow cytometry plots of endogenous and transferred CD8$^+$ T cells (gated TCRβ$^+$CD8$^+$ cells) from splenocytes from OT-I *Rag2*$^{-/-}$ aged mice 3 days after SNI with control IgG or CXCL13 mAb. (**J**) Percentage of transferred CD8$^+$ T cells from spleen (n = 3 or 4). (**K**) Plots showing wild-type or *Cxcr5*$^{-/-}$ CD8$^+$ T cells after sorting. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 [one-way ANOVA with post hoc Tukey's test in (B), (C), (E), (G), (H), and (J)]. Data are means ± SEM and represent four experiments.

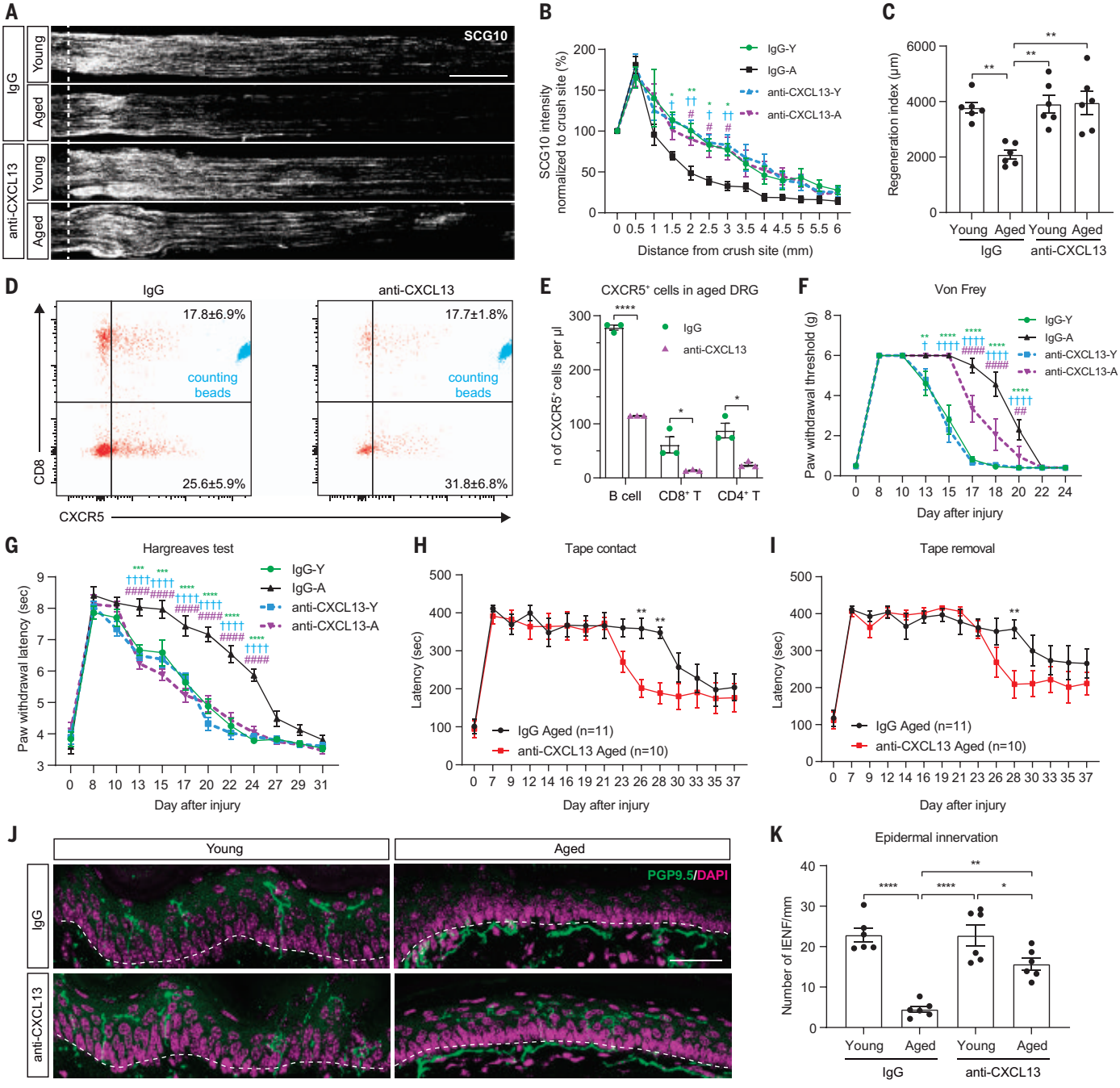


Fig. 7. CXCL13 neutralization enhances axonal regeneration, epidermal reinnervation, and sensory functional recovery after SNI in aged mice. (**A**) Representative SCG10 immunostaining of sciatic nerves after control IgG or CXCL13 mAb treatment 3 days after SNI. The dashed line indicates the proximal injury site. Scale bar, 500 μm. (**B**) Normalized SCG10 intensity (n = 6, *P < 0.05, **P < 0.01, IgG-Y versus IgG-A; †P < 0.05, ††P < 0.01, anti-CXCL13-Y versus IgG-A; #P < 0.05, anti-CXCL13-A versus IgG-A). (**C**) Analysis of regeneration index (n = 6). (**D**) Representative flow cytometry plots of CXCR5$^+$ and CD8$^+$ T cells (gated CD45$^+$TCRβ$^+$CD19$^-$ cells) from the sciatic DRG of aged mice 3 days after SNI. (**E**) Quantification of CXCR5$^+$ B, CD8$^+$, and CD4$^+$ T cell number from DRG 3 days after SNI (n = 3). (**F** and **G**) Mechanical sensitivity by von Frey test (F) and thermosensitivity analysis by Hargreaves test (G) after SNI. **P < 0.01, ***P < 0.001, ****P < 0.0001 (IgG-Y versus IgG-A); †P < 0.05, ††††P < 0.0001 (anti-CXCL13-Y versus IgG-A); ##P < 0.01, ####P < 0.0001 (anti-CXCL13-A versus IgG-A) (n = 10 or 11; two-way ANOVA with post hoc Sidak's test). (**H** and **I**) Sensory responses by adhesive tape contact (H) and removal (I) test after SNI (n = 10 or 11). (**J**) Representative coimmunostaining of PGP9.5 and DAPI shows the epidermal innervation of the hindpaw interdigital skin 18 days after SNI. The dashed lines indicate the boundary of the epidermis and dermis. Scale bar, 50 μm. (**K**) Quantification of the number of intra-epidermal nerve fibers (IENF) per mm of interdigital skin (n = 6). *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 [two-way ANOVA with post hoc Sidak's test in (B), (F), (G), (H), and (I); one-way ANOVA with post hoc Tukey's test in (C) and (K); unpaired Student's t test in (E)]. Data are means ± SEM and represent three [(B), (C), and (E)] or two [(F), (G), (H), (I), and (K)] experiments.

enhanced axonal regeneration and sensory recovery.

CXCL13 is typically expressed in lymphoid organs to control the recruitment of B cells and antigen-presenting cells (*20*). The CXCL13 receptor CXCR5 is highly expressed by a subset of B cells and by $CD4^+$ T cells that are attracted to CXCL13-rich sites such as the lymphoid tissue. CXCL13 has also been found to be expressed by astrocytes and DRG neurons regulating pain (*21*, *22*). $CXCR5^+CD8^+$ T cells have been described as follicular effector memory T cells and have been implicated in the orchestration of the follicular T and B cell responses to infection (*23*, *24*). Clonally expanded $CD8^+$ T cells are present in neurogenic niches in the aged mouse brain whose activity is required to inhibit neural stem cell proliferation (*25*). However, the role of $CD8^+$ T cells in the nervous system can be either neurotoxic or neuroprotective, depending on the specific experimental conditions (*26*–*28*).

We also observed that $CXCR5^+$ B cells are attracted by CXCL13 gradients toward the DRG in aged animals and after CXCL13 overexpression, but remain trapped in the perivenular space. Unlike $CXCR5^+CD8^+$ T cells, $CXCR5^+$ B cells do not migrate into the parenchyma and do not affect neuronal regenerative ability. Whether CXCL13 and $CXCR5^+$ B and T cells populate the perivenular spaces or the parenchyma in the aged brain and whether they may contribute to tissue damage remains unexplored.

$CXCR5^+CD8^+$ T cells are recruited and progressively accumulate into the DRG, where they reside until they act as effector T cells upon MHC I antigen presentation by injured neurons. Although the nature of the aging and injury-dependent antigens remains elusive, we speculate that the combination of injury-dependent antigens with viral peptides accumulated and processed during the lifespan may lead to T cell–dependent neuronal MHC I recognition and inhibition of axonal regeneration after injury.

Neuronal MHC I has been implicated in synaptic maintenance of sciatic motor neurons and locomotor function after sciatic nerve or spinal cord injury in young animals (*29*, *30*). However, its role in axonal regeneration of aged sensory neurons had not been explored. Here, we have shown that the inhibitory signaling limiting axonal regeneration in aged DRG neurons appears to rely on neuronal MHC I–$CD8^+$ T cell interactions that elicit intracellular signals such as perforin, granzyme B, and caspase 3 activation (*31*–*33*). Active caspase 3 is needed to reduce pAKT, likely by AKT cleavage (*15*), resulting in decrease of downstream signals. Depletion of $CD8^+$ T cells or inhibition of caspase 3 activation counteracts aging-dependent regenerative decline and results in a significant increase in both pAKT and pS6. These are strongly developmentally regulated signals within the regenerative PI3K and mTOR pathways (*17*). Thus, the induction of these pathways in aged DRG neurons is likely to be crucial to promote regenerative signaling and merits further exploration.

In line with our data, CXCL13 expression was previously shown to be reduced in the spleens and lymph nodes of aged mice relative to young mice (*34*, *35*), leading to reduced accumulation of B and T cells. The CXCL13-dependent shift of B cells and T cells from secondary lymphoid organs to peripheral tissues may be a general phenomenon leading to tissue damage beyond the peripheral nervous system.

The delivery of CXCL13 mAbs that promote axonal regeneration, skin reinnervation, and recovery of neurological function represents a translational opportunity for improving repair after injury. CXCL13 neutralization may reduce $CXCR5^+$ follicular helper B cell and T cell homing, with a risk of increased susceptibility to viral infection. However, the CXCL13 mAb used here is currently in preclinical development and has been effective in ameliorating pathology and disability in animal models of rheumatoid arthritis and multiple sclerosis without obvious toxicity (*36*). Our results therefore suggest that CXCL13 neutralization may be considered for the treatment of axonal injuries affecting the elderly. More generally, they describe aging-specific regenerative mechanisms, suggesting the need for targeted interventions that consider age at the time of injury or disease initiation.

Materials and methods

Mice

Wild-type 8- to 10-week-old C57BL/6J young mice were obtained from Charles River, whereas 20- to 22-month-old aged mice were supplied by either Charles River or by M. Jucker (Hertie Institute for Clinical Brain Research, University Hospital Tübingen, Germany). OT-I *Rag2*$^{-/-}$ mice on a C57BL/6 genetic background were provided by M. Botto (Imperial College London). *Cxcr5*$^{-/-}$ mice on a C57BL/6 genetic background were donated by T. Arnon (University of Oxford). All animal procedures were performed in accordance with the UK Animals Scientific Procedures Act (1986) and approved by the ethical committee of Imperial College London.

DRG cell culture

Sterile 15-mm glass coverslips in 24-well plates were coated with 0.003% poly-L-ornithine (Sigma) diluted in H_2O for 1 hour at 37°C. After three washes with water, the coverslips were coated with laminin (1 μg/ml; Millipore) prepared in PBS (Invitrogen) for 3 hours at room temperature. After two washes with PBS, the wells were filled with DMEM and placed at 37°C until cells were plated. DRGs were dissected and collected in Hank's balanced salt solution (HBSS; Invitrogen) on ice. HBSS was then removed and replaced with 500 μl of digest solution [Dispase II (5 mg/ml, Sigma) and 2.5 mg/ml collagenase type II (2.5 mg/ml, Worthington] in DMEM (Invitrogen) for 40 min in a 37°C water bath with gentle tube flicks every 5 min. DRG were collected by centrifugation at 72*g* for 2 min after digestion and washed once with warm DRG medium [DMEM/F12 (Invitrogen), 1× B27 (Invitrogen), and 10% fetal bovine serum (FBS; Sigma)]. DRG were then resuspended in 1 ml of warm DRG medium and dissociated by pipetting with fire-polished Sigmacote (Sigma)–coated glass pipette 15 times. Undissociated tissue was filtered with a 100-μm cell strainer (BD Falcon) and cells were counted under a bright-field microscope with a hemocytometer. After cell counting, cells were centrifuged, resuspended with warm DRG culture medium [DMEM/F12 (Invitrogen), 1× B27 (Invitrogen), and 1% penicillin and streptomycin (Invitrogen)] and plated (4000 cells per well). The DRG cells were cultured at 37°C in 5% CO_2 for 24 hours.

DRG neurite outgrowth analysis

The average neurite length of cultured DRG cells was measured by using Neurolucida software (MBF Bioscience) with at least 100 cells (five fields per coverslip randomly selected) per condition in technical and biological triplicate.

Immunocytochemistry

Cultured DRG cells were fixed with ice-cold 4% paraformaldehyde (PFA; Sigma) in 1× PBS for 30 min on ice and washed with 1× PBS three times. Cells were blocked in 1× PBS, 0.1% Triton X-100 (Sigma), and 5% normal goat serum (Abcam) at room temperature for 1 hour followed by the incubation with primary antibodies in 1× PBS, 0.1% Triton X-100, and 2% normal goat serum overnight at 4°C. Cells were then washed three times with 1× PBS and incubated with secondary antibodies in 1× PBS, 0.1% Triton X-100, and 2% normal goat serum for 2 hours at room temperature in the dark. Cells were washed three times with 1× PBS and incubated with 4,6-diamidino-2-phenylindole (DAPI, ThermoFisher, 62248, 0.2 μg/ml) in 1× PBS, 0.1% Triton X-100 for 10 min followed by two washes with 1× PBS. The coverslips with cells facing down were transferred and mounted with Antifade mounting medium (Vectashield, H-1000) on the slides. Detailed antibody information is provided in table S1.

Immunohistochemistry

Mice were anesthetized by ketamine-xylazine mixture (80 mg of ketamine and 10 mg of xylazine per kilogram of body weight) via intraperitoneal (i.p.) injection and transcardially perfused with 20 ml of ice-cold PBS

followed by 4% PFA in a flow rate of 5 ml/min. DRG, sciatic nerves, or glabrous skin of hindpaws were excised and post-fixed in 4% PFA for 2 hours on ice and transferred to 30% sucrose (Sigma) for 3 days prior to embedding in OCT compound (Tissue-Tek). DRG were sectioned at 10 μm, sciatic nerves at 12 μm, and interdigital footpads at 30 μm perpendicularly to the skin surface with a cryostat (Leica). Sections were treated with a blocking solution containing 10% normal goat serum or normal donkey serum (Abcam) and 0.3% Triton X-100 in PBS for 1 hour at room temperature and stained with primary antibodies at 4°C overnight. Sections were washed with 1× PBS three times followed by secondary antibody incubation at room temperature for 2 hours. Sections were then stained with DAPI (ThermoFisher, 62248, 0.2 μg/ml) for 15 min at room temperature followed by three washes with 1× PBS and then mounted with Antifade mounting medium (Vectashield, H-1000). DRG were imaged with a Nikon Eclipse TE2000 microscope using 20× magnification (DIC Plan-Apochromat VC objective and QImaging OptiMOS detection system). Sciatic nerves (20× magnification, HC Plan-Apochromat CS2) and epidermis (40× magnification with oil lens, HC Plan-Apochromat CS2) were imaged with a Leica TCS SP8 confocal laser scanning microscope in Leica DFC9000 GTC detection system. Detailed antibody information is provided in table S1.

Sciatic nerve crush injury

Sciatic nerve was crushed following the protocol previously described (*37*). Mice were anesthetized with 2% isoflurane in oxygen (1 liter/min) and injected with buprenorphine (0.1 mg/kg body weight) and rimadyl (5 mg/kg body weight) subcutaneously (s.c.) for analgesia. An incision across the mid-thigh was made in the skin, the sciatic nerve was exposed by opening the fascial plane, and a small deeper incision was made between the gluteus maximus and the anterior head of the biceps femoris. The proximal side of exposed nerve where the three fascicles are sequentially aligned was placed on the bottom jaw of a super-fine hemostatic forceps (Fine Science Tools, 13020-12) 1.5 mm from their tip. The nerve was crushed for 30 s at three clicks of the forceps with no nerve stretch. The skin was then closed with suture clips. Mice were administered buprenorphine (0.1 mg/kg body weight) twice per day and rimadyl (5 mg/kg body weight) daily by s.c. injection for 3 days post-operation. Sham surgery was performed to expose the nerves without injury.

Retrograde tracing after nerve injury

Mice were anesthetized following the protocol of sciatic nerve crush injury. The skin and muscle at the distal thigh were opened to expose the nerve and 1 μl of 1% Alexa Fluor 555-conjugated cholera toxin subunit B (CTB, ThermoFisher) diluted in sterile saline was injected in the nerve 8 mm distal to the injury site with Hamilton syringe and needle (NDL small RN ga34/15mm/pst45°) (Hamilton) as described (*38*). The skin was closed with suture clips.

RNA preparation and sequencing

RNA sequencing was performed using RNA from the whole sciatic DRG tissue of young or aged mice 24 hours after sham or sciatic nerve injury. Specifically, sciatic DRG were dissected from one mouse per sample and stored in RNAlater stabilization solution (ThermoFisher). Tissue was homogenized with RNase free micro pestle and RNA was extracted using RNeasy kit (Qiagen) following the manufacturer's protocol. On-column DNase digestion was performed with the RNase-free DNase set (Qiagen) for 15 min to remove DNA. RNA concentrations and purity were measured using Agilent 2100 Bioanalyzer (Agilent). RNA with RNA integrity number (RIN) above 7.5 was used for library preparation. Libraries were prepared at Ospedale San Raffaele (Milan) using the TruSeq mRNA Sample Preparation kit (Illumina) and sequenced using Illumina HiSeq 2500 100-cycle, paired-end sequencing.

Bioinformatic analysis

RNA sequencing data analysis was performed as published (*39*). FDR-corrected *P* value was implemented with Benjamini-Hochberg (BH) correction. DE genes (FDR < 0.05) of three comparisons (SNI Young, Sham Aged, and SNI Aged versus Sham Young, respectively) were analyzed for Gene Ontology (GO) and KEGG pathways using DAVID v6.7 (https://david-d.ncifcrf.gov/) with all expressed genes in our dataset as the background. The categories of GO and KEGG pathways were selected by $P < 0.01$ for both up- and down-regulation. Area proportional Venn diagrams were generated using Biovenn (www.biovenn.nl/). The differentially up-regulated genes from SNI Aged versus Sham Young involved in immune response collected from GO and KEGG pathways were analyzed for protein-protein network established by STRING (https://string-db.org/). The stringent criteria were selected using the active interaction sources of experiments and databases, as well as the highest confidence of the minimum required interaction score. The network was visualized by Cytoscape v3.7.2 (https://cytoscape.org/). For the generation of Circos plots of VJ segment usage, we used MiXCR (v3.0.13) to identify T cell receptor clonotypes from whole DRG RNA-seq data. MiXCR's "analyze shotgun" method was run directly on the fastq files from the RNA-seq experiment using default parameters for mouse RNA to identify clonotypes. The T cell receptor clonotypes were then pooled among replicates and imported to VDJtools v1.2.1. The "PlotFancyVJUsage" method was used to generate the Circos plot separately for each biological condition, using default parameters.

CXCL13 ELISA assay

Sciatic DRG dissected from two young or aged mice per sample 3 days after sham or sciatic nerve injury were collected and homogenized in 150 μl Assay Diluent A supplied from Mouse CXCL13 ELISA kit (ThermoFisher) with protease inhibitors (Roche) to produce the data in Fig. 1I. Total DRG and spleen were dissected from one young or aged OT-I *Rag2*$^{-/-}$ mouse per sample and homogenized in 150 μl and 2 ml of Assay Diluent A buffer with protease inhibitor, respectively, to produce the data in Fig. 5C. After homogenization, Triton X-100 was added to a final 1% concentration. Samples were frozen in liquid nitrogen and thawed followed by centrifugation at 10,000*g* for 5 min to remove the debris. Protein concentration was quantified using BCA protein assay kit (ThermoFisher) and 100 μl of cell lysate was used for the CXCL13 ELISA assay using Mouse CXCL13 ELISA kit according to the manufacturer's protocols. CXCL13 ELISA measurements were obtained by measuring absorbance on an ELISA plate reader set at 450 nm and subtracting 550-nm values to correct for optical imperfections in the microplate. CXCL13 concentrations in the cell lysate were calculated according to the standard curve.

Quantification of immunostaining

Arbitrary fluorescence intensity was measured with ImageJ software (NIH) and normalized fluorescence intensity was calculated by subtracting the intensity of the background immunofluorescent signal. The relative fold change was calculated by comparing the normalized fluorescence intensity of the experimental group versus the control group. The cells designated positive for the expression of cleaved caspase 3, perforin, and granzyme B showed fluorescence intensity at least twofold above background. Longitudinal sections of sciatic nerve stained with SCG10 were analyzed for axonal regeneration. SCG10 arbitrary intensity was measured along the nerve and percentage of intensity in every 0.5 mm was quantified away from the proximal site of crush. A regeneration index was evaluated by calculating the distance away from the proximal injury site where the intensity was reduced to 50% level with respect to the proximal site. For epidermal and dermal reinnervation, the number of IENF indicated by PGP9.5 staining from sagittal sections of the skin was measured per unit of volume. Animals were randomly and equally grouped.

Immunoblotting

Sciatic DRG were dissected and collected in 1× PBS with protease inhibitor (Roche) and

phosphatase inhibitor (Roche). Tissues were then homogenized, and protein lysate was extracted by RIPA buffer. Protein concentration was measured with BCA protein assay kit (ThermoFisher) and 20 μg of protein was loaded to 10% SDS-PAGE gel and transferred on nitrocellulose membrane of iBlot Transfer Stack (ThermoFisher) by iBlot 2 Gel Transfer Device (ThermoFisher) under the default P0 program. The membrane was blocked with 5% milk in TBST buffer for 1 hour at room temperature (RT) and incubated with primary antibodies at 4°C overnight. The membranes were then washed with TBST and incubated with horseradish peroxidase (HRP)–conjugated secondary antibodies at RT for 1 hour followed by chemiluminescent detection with Amersham Hyperfilm (GE Healthcare Life Sciences) after incubation with ECL substrate (ThermoFisher). Detailed antibody information is provided in table S1.

Flow cytometry

For cell characterization in vivo, young or aged naïve or sciatic nerve-injured animals were injected with 3 μg of APC-conjugated anti-CD45 (I3/2.3, Biolegend) intravenously (i.v.) for 3 min. Peripheral blood or DRG were collected in 10 ml of 1× PBS with 2 mM EDTA and RPMI-1640 medium (ThermoFisher), respectively, after anesthesia. Blood cells were collected by centrifugation at 300*g* for 5 min at RT and erythrocytes were lysed using 1× RBC lysis buffer (Biolegend). Cells were washed with 1× PBS three times and filtered by 70-μm cell strainer (Corning) to remove the clumps and resuspended by fresh FACS buffer (0.5% BSA and 2 mM EDTA in PBS) for staining. DRG were washed by 1× PBS once and dissociated with digestion buffer [CLSPA (0.5 mg/ml, Worthington, #LS005273)] and DNase I (7.5 μg/ml, Roche, #10104159001) in RPMI-1640 at 37°C for 30 min. After digestion, DRG tissues were dispersed and filtered through a 70-μm cell strainer to collect the single-cell suspension and washed with rinse buffer (5% FBS in RPMI-1640). After centrifugation, the pellet was resuspended with 1 ml of DNase solution [250 μg/ml of DNase I in 1× DNase buffer (1.21 mg/ml of Tris Base, 0.5 mg/ml of $MgCl_2$, and 73 μg/ml of $CaCl_2$ in RPMI-1640)] at RT for 30 min. Cells were washed with FACS buffer for staining.

Isolated cells from the blood and DRG samples were treated with 1.0 μg of TruStain FcX anti-mouse CD16/32 antibody (Biolegend) per 10^6 cells in a 100-μl volume to block Fc receptors for 10 min on ice and stained with primary antibodies in FACS buffer with Brilliant Stain Buffer (BD Horizon). LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (ThermoFisher) was also used to exclude the dead cells from staining. Cells were stained with the cocktail antibodies at 4°C in the dark for 20 min and stained with PE-streptavidin (2 μg/ml, Biolegend) at 4°C for 20 min after washing. Fluorescence Minus One (FMO) and PE-streptavidin staining controls were used for the gating boundaries. To minimize anti-CXCR5 shedding before cell acquisition, stained cells were immediately fixed with IC fixation buffer (ThermoFisher) for 20 min at RT. Cells were washed with 1× PBS and resuspended with 180 μl of FACS buffer. Precision Counting Beads (20 μl; Biolegend, 424902) were added to each sample before cell acquisition by flow cytometer LSR II (BD Biosciences). All DRG cells and 5000 blood cells or splenocytes were acquired for quantification. Detailed antibody information is provided in table S1.

NF-κB and Cxcl13 expression in cultured DRG neurons

Isolated DRG primary cells were plated with cytokines and incubated at 37°C in 5% CO_2 for 24 hours. The following compounds were used in cell culture: recombinant mouse TNF-α (410-MT, R&D Systems, 500 pg/ml), recombinant mouse IL-1β (401-ML, R&D Systems, 600 pg/ml), recombinant mouse BAFF (8876-BF, R&D Systems, 100 ng/ml), and recombinant mouse lymphotoxin α1/β2 (9968-LY, R&D Systems, 100 ng/ml). Selective IKK inhibitor (10 μM; PS1145, TOCRIS, #4569) was in turn used with or without lymphotoxin α1/β2 stimulation for 24 hours. Total RNA from cultured cells was extracted using RNeasy kit (Qiagen) according to manufacturer's guidelines. cDNA was then synthesized with SuperScript II Reverse Transcriptase (ThermoFisher). qRT-PCR was performed using KAPA SYBR FAST qPCR kit (Sigma, KK4601) from Applied Biosystems 7900 Real-Time PCR System. Ct values normalized versus *Gapdh* as a housekeeping gene were exported and relative fold change of gene expression compared to vehicle control was calculated by $2^{-\Delta\Delta Ct}$. Primers used in qPCR were as follows: *Nfκb1* forward 5′-GAAATTCCTGATCCAGACAAAAAC-3′ and reverse 5′-ATCACTTCAATGGCCTCTGTGTAG-3′, *Nfκb2* forward 5′-GGCCGGAAGACCTATCCTACT-3′ and reverse 5′-CTACAGACACAGCGCACACT-3′, *Cxcl13* forward 5′-CTCTCCAGGCCACGGTATT-3′ and reverse 5′-TAACCATTTGGCACGAGGAT-3′, *Gapdh* forward 5′-TCAACAGCAACTCCCACTCTTCCA-3′ and reverse 5′-ACCCTGTTGCTGTAGCCGTATTCA-3′.

Cell migration in culture after CXCL13 overexpression

AAV5-GFP (Vector Biolabs, #7073) or AAV5-*Cxcl13*-GFP (AAV-256398) was added (1 μl per well) to DRG primary cell culture (8000 cells per well in 24-well plates) for a 5-day incubation. Thereafter, culture medium was collected and centrifugated at 300*g* for 5 min and 200 μl of culture medium was transferred to each receiving well of Corning HTS Transwell 96-well permeable supports, 5.0-μm pore size (Sigma, CLS3387). A fresh wild-type mouse spleen was meshed on 70-μm cell strainer and collected cells were treated with RBC lysis buffer to remove the erythrocytes. Purified splenocytes after PBS washing were incubated with 250 μg/ml of DNase I for 30 min at RT. After DNA digestion and PBS washing, cell numbers were counted and adjusted to the density of 5×10^5 cells per 100 μl by DRG culture medium. Cell suspension (100 μl per well) was carefully loaded to the transwell, which was then placed onto the receiving plate to make sure the liquid from each well was in contact with the transwell. The plate was incubated at 37°C for 4 hours. After migration, the transwell was removed and the medium in the receiving plate was collected. The medium was pooled from six wells as one sample. Cells were stained and acquired by flow cytometer and three replicates per group were measured for quantification.

To validate the successful overexpression or secretion of CXCL13 in cell culture, total RNA was extracted for cDNA synthesis followed by *Cxcl13* qRT-PCR and DNA electrophoresis, or culture medium was collected for CXCL13 assay using mouse CXCL13 ELISA kit as described above.

In vivo DRG neuron CXCL13 overexpression

Eight-week-old mice were anesthetized with 2% isoflurane in 1 liter/min oxygen and administered with buprenorphine (0.1 mg/kg body weight) and rimadyl (5 mg/kg body weight) by s.c. injection. The skin and muscle at the distal thigh were opened to expose the nerve, and 2 μl of AAV-GFP or AAV-CXCL13-GFP were injected in the nerve bilaterally at the distal site of the thigh with Hamilton syringe and needle (NDL small RN ga34/15mm/pst45°) (Hamilton). The skin was closed with suture clips. Animals were monitored for 6 weeks after CXCL13 overexpression. Recombinant mouse IFN-γ (10 μg, Biolegend, #575308) and 25% D-mannitol (200 μl, Sigma, M4125) were i.p. and i.v. injected, respectively, into each mouse 2 days before sciatic nerve crush. Sciatic DRG were collected for tissue dissociation with CLSPA and cells were then stained with LIVE/DEAD-Aqua and cell markers for flow cytometry. Sciatic DRG and nerves were collected and fixed for the measurement of gene overexpression and axonal regeneration.

Cell depletion experiments

To deplete $CD8^+$ or $CD4^+$ T cells in vivo, 22- to 24-month-old mice were i.p. injected with 200 μg of anti-mouse CD8α (YTS 169.4, BioXcell, BE0117) or anti-mouse CD4 (YTS 191, BioXcell, BE0119). An equivalent amount of rat IgG2b (LTF-2, BioXcell, BE0090) was injected into the control group of animals. Depletion of B cells was performed following a

strategy described previously (*40*). Aged mice were i.p. injected with antibody cocktail consisting of 150 μg of InvivoMAb anti-mouse CD19 (1D3, BioXcell, BE0150), 150 μg of InvivoMAb anti-mouse B220 (RA3.3A1/6.1, BioXcell, BE0067), and 150 μg of InvivoMAb anti-mouse CD22 (Cy34.1, BioXcell, BE0011). Forty-eight hours later, the mice were injected with a secondary antibody InvivoMAb anti-rat kappa (MAR18.5, BioXcell, BE0122), at a dose of 150 μg per mouse. Rat IgG2a (2A3, BioXcell, BE0089), polyclonal rat IgG (BioXcell, BE0094), mouse IgG1 (MOPC-21, BioXcell, BE0083), and mouse IgG2a (C1.18.4, BioXcell, BE0085) isotype controls were injected in the same dosage. All injections were performed twice 1 week prior to the sciatic nerve injury and once more the day after injury.

DRG neuron and OT-I CD8$^+$ T cell coculture

Eight-week-old wild-type mice were injected with PBS or 10 μg of recombinant mouse IFN-γ 2 days before being killed for DRG dissection and DRG neuron culture. Naïve OT-I CD8$^+$ T cells were purified from 8-week-old OT-I mice using the MojoSort Mouse CD8 T Cell Isolation kit (Biolegend). Purified live CD8$^+$ T cells were counted by Trypan Blue (ThermoFisher) staining. Approximately 4000 DRG neurons per well were plated with 10,000 OT-I CD8$^+$ T cells with or without $OVA_{257–264}$ peptide (100 nM, SIINFEKL, InvivoGen) at 37°C for 24 hours. The culture medium was then removed and the cells were fixed with 4% PFA. This was followed by two washes with 1× PBS for immunostaining.

GAr-rtTA–mediated immune evasion

AAV5-rtTA or AAV5-GAr-rtTA viral particles, which were provided by J. Verhaagen (Netherlands Institute for Neuroscience), were mixed with the viral particles expressing doxycycline-inducible luciferase (AAV5-dox-i-luciferase) in a 1:1 ratio to be titer-matched by PBS+5% sucrose dilution. Viral mix (3 μl) was injected into the sciatic nerve of aged mice bilaterally. Five weeks later, doxycycline (Stemcell, #72742) reconstituted in PBS was delivered by daily i.p. injection at a dosage of 50 mg/kg body weight for 1 week before sciatic nerve injury and 3 days after injury. Upon killing, sciatic DRG and nerves were dissected for immunostaining.

CXCL13 neutralization

CXCL13 mAb (5378) and control IgG2a mAb (2510) were kindly supplied by Vaccinex Inc. (Rochester, NY). For DRG ex vivo culture, young and aged mice were treated with control IgG or CXCL13 mAb (30 mg/kg body weight, i.p.) three times 1 week before DRG dissection and primary cell culture. Isolated DRG cells were incubated at 37°C in 5% CO_2 for 24 hours followed by neurite outgrowth measurement. For CXCL13 neutralization in vivo, control IgG or CXCL13 mAb (30 mg/kg body weight) was i.p. injected into the young or aged mice 4 hours after sciatic nerve crush and daily until killing on day 3. For the chronic nerve regeneration study, 30 mg/kg body weight control IgG or CXCL13 mAb was i.p. injected three times per week until the animals were killed beginning 4 hours after injury.

Adoptive cell transfer

Spleens from wild-type or *Cxcr5*$^{-/-}$ mice were harvested. Splenocytes were extracted after red blood cell lysis using RBC lysis buffer (Biolegend, 420302). CD8$^+$ T cells were then purified with mouse CD8$^+$ T cell isolation kit (Miltenyi Biotec, 130-104-075) and LS columns (Miltenyi Biotec, 130-042-401). Wild-type CD8$^+$ T cells or *Cxcr5*$^{-/-}$ CD8$^+$ T cells (2×10^7) were i.v. injected into each aged OT-I *Rag2*$^{-/-}$ mouse (20 to 22 months old) and a sciatic nerve crush was performed 3 days after cell transfer. Control IgG or CXCL13 mAb (30 mg/kg body weight) was i.p. injected into the recipients immediately after cell transfer and daily until killing on day 3 after nerve injury. For wild-type CXCR5$^+$CD8$^+$ T cell transfer, purified CD8$^+$ T cells from spleen and lymph nodes of adult wild-type mice (4 to 6 months old) were stained with LIVE/DEAD, anti-CD8, anti-TCRβ, biotin-conjugated anti-CXCR5 antibodies for 20 min followed by washing and staining with APC-streptavidin. Live TCRβ$^+$CD8$^+$CXCR5$^+$ cells were sorted with BD FACSAria III flow cytometer. Cells (5×10^6, sorted from spleen and lymph nodes of ~10 donors) per recipient were adoptively transferred to the young (8 to 10 weeks old) and aged OT-I *Rag2*$^{-/-}$ mice (20 to 22 months old). Control groups were injected with an equivalent amount of PBS. Three days later, a sciatic nerve crush was performed, and the animals were killed 3 days after nerve injury. DRG, spleen, and blood samples were then collected after a 3-min labeling of blood-exposed cells by i.v. injection of PE-Cy7–conjugated anti-CD45. Isolated DRG cells, splenocytes, and blood cells were stained with LIVE/DEAD, primary antibodies, and tetramer (SIINFEKL) for 30 min at room temperature. The cells were washed and stained with APC-streptavidin (2 μg/ml) for 20 min at 4°C. Cell samples were acquired by flow cytometer LSR II (BD Biosciences). Detailed antibody information is provided in table S1.

In vivo administration of IKK and caspase 3 inhibitors

PS1145 (TOCRIS, #4569) was reconstituted with 0.5% methyl cellulose (Sigma), 0.2% Tween-80 (Sigma) in water. The suspension was vortexed for 2 min and sonicated at 37°C for 20 min (LabSonic, B. Braun Biotech International; 0.5-mm-diameter sonotrode, three times per 5 s, amplitude 80%). The methyl cellulose solution or PS1145 suspension was delivered at 50 mg/kg body weight by intragastric gavage immediately after sciatic nerve crush and at day 2. Specific inhibition of caspase 3 in vivo was performed by i.v. delivery of Z-DEVD-FMK (Santa Cruz, #210344-95-9) in 1% DMSO at a dosage of 20 mg/kg body weight daily after sciatic nerve injury.

Behavioral assessment

Mice with sciatic nerve crush and chronic administration (i.p. injection) of control IgG or CXCL13 mAb (30 mg/kg body weight, three times per week) were used for behavioral assessments. The left side of sciatic nerve and hindpaw were injured and behaviorally assessed at day 0 and from day 7 to day 42 followed by the tissue collection at day 43. Animals were randomly and equally grouped. All investigators participating in behavioral assessments were blind to the groups and the treatments (*41*).

Mechanical sensitivity test

Mice were placed in the test chamber for 30 min to acclimate before initiating the behavioral assessment. Mechanical sensitivity was determined by probing the plantar surface of the left hindpaw with calibrated Von Frey filaments ranging from 0.4 to 4 g. To give the sensory receptors enough time to return to baseline, the interval time was at least 30 s between each trial. A quick hindpaw withdrawal was considered a positive response. Five trials were conducted for each mouse. The withdrawal threshold to the mechanical stimulation was defined as the lowest force that resulted in at least three withdrawals in five trials, as described (*42*).

Thermal sensitivity test

For thermal sensitivity, the Hargreaves test (Ugo Basile, catalog number 37370) was performed as described (*43*, *44*). Briefly, mice were placed on a glass floor and separated by plastic chambers. Mice were given 30 min to acclimate. The thermal heat stimuli (infrared radiation source, intensity = 50) were carefully placed under the plantar surface of the hindpaw for no more than 10 s (*44*). The hindpaw withdrawal time was recorded automatically. Five trials were conducted on each paw and the average was calculated including the longest and shortest withdrawal time.

Adhesive removal

Small adhesive stimuli (6-mm round adhesive labels) were placed on the left hindpaw of the mouse. The time to make first contact with both forepaws or mouth, as well as the time to completely remove the adhesive, were then recorded. Daily 10-min training sessions per mouse were performed for 1 week prior to the nerve injury. Each mouse was subjected to three trials. All testing was performed in the

animal's home cage by a blind investigator as described (*39*, *45*).

Statistical analysis

Data were statistically analyzed with Graphpad Prism 9.3 (Graphpad Software Inc., La Jolla, CA) and expressed as mean ± SEM unless otherwise stated. Statistical comparisons were analyzed using unpaired two-tailed Student's *t* test for two groups. Multiple groups were analyzed using one-way analysis of variance (ANOVA) with post hoc Tukey's correction or two-way ANOVA with post hoc Sidak's test. Statistical significance was considered as $P < 0.05$. All experiments were replicated two or three times. The *n* defining biological replicates was always determined by power analysis with power set at 80% and significance level at 5%.

REFERENCES AND NOTES

1. M. J. DeVivo, Y. Chen, Trends in new injuries, prevalent cases, and aging with spinal cord injury. *Arch. Phys. Med. Rehabil.* **92**, 332–338 (2011). doi: 10.1016/j.apmr.2010.08.031; pmid: 21353817
2. A. Pestronk, D. B. Drachman, J. W. Griffin, Effects of aging on nerve sprouting and regeneration. *Exp. Neurol.* **70**, 65–82 (1980). doi: 10.1016/0014-4886(80)90006-0; pmid: 7418774
3. D. W. Vaughan, Effects of advancing age on peripheral nerve regeneration. *J. Comp. Neurol.* **323**, 219–237 (1992). doi: 10.1002/cne.903230207; pmid: 1401257
4. E. Verdú, D. Ceballos, J. J. Vilches, X. Navarro, Influence of aging on peripheral nerve function and regeneration. *J. Peripher. Nerv. Syst.* **5**, 191–208 (2008). doi: 10.1111/j.1529-8027.2000.00026.x; pmid: 11151980
5. H. Kang, J. W. Lichtman, Motor axon regeneration and muscle reinnervation in young adult and aged animals. *J. Neurosci.* **33**, 19480–19491 (2013). doi: 10.1523/JNEUROSCI.4067-13.2013; pmid: 24336714
6. M. W. Painter *et al*., Diminished Schwann cell repair responses underlie age-associated impaired axonal regeneration. *Neuron* **83**, 331–343 (2014). doi: 10.1016/j.neuron.2014.06.016; pmid: 25033179
7. D. Weiskopf, B. Weinberger, B. Grubeck-Loebenstein, The aging of the immune system. *Transpl. Int.* **22**, 1041–1050 (2009). doi: 10.1111/j.1432-2277.2009.00927.x; pmid: 19624493
8. N. Barzilai, D. M. Huffman, R. H. Muzumdar, A. Bartke, The critical role of metabolic pathways in aging. *Diabetes* **61**, 1315–1322 (2012). doi: 10.2337/db11-1300; pmid: 22618766
9. L. C. D. Pomatto, K. J. A. Davies, The role of declining adaptive homeostasis in ageing. *J. Physiol.* **595**, 7275–7309 (2017). doi: 10.1113/JP275072; pmid: 29028112
10. J. C. Lo *et al*., Coordination between NF-κB family members p50 and p52 is essential for mediating LTβR signals in the development and organization of secondary lymphoid tissues. *Blood* **107**, 1048–1055 (2006). doi: 10.1182/blood-2005-06-2452; pmid: 16195333
11. S. A. van de Pavert *et al*., Chemokine CXCL13 is essential for lymph node initiation and is induced by retinoic acid and neuronal stimulation. *Nat. Immunol.* **10**, 1193–1199 (2009). doi: 10.1038/ni.1789; pmid: 19783990
12. E. Dejardin *et al*., The lymphotoxin-β receptor induces different patterns of gene expression via two NF-κB pathways. *Immunity* **17**, 525–535 (2002). doi: 10.1016/S1074-7613(02)00423-5; pmid: 12387745
13. J. Levitskaya *et al*., Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. *Nature* **375**, 685–688 (1995). doi: 10.1038/375685a0; pmid: 7540727
14. E. R. Burnside *et al*., Immune-evasive gene switch enables regulated delivery of chondroitinase after spinal cord injury. *Brain* **141**, 2362–2381 (2018). doi: 10.1093/brain/awy158; pmid: 29912283
15. J. Chu, E. Lauretti, D. Praticò, Caspase-3-dependent cleavage of Akt modulates tau phosphorylation via GSK3β kinase: Implications for Alzheimer's disease. *Mol. Psychiatry* **22**, 1002–1008 (2017). doi: 10.1038/mp.2016.214; pmid: 28138159
16. K. Liu *et al*., PTEN deletion enhances the regenerative ability of adult corticospinal neurons. *Nat. Neurosci.* **13**, 1075–1081 (2010). doi: 10.1038/nn.2603; pmid: 20694004
17. W. Chen *et al*., Rapamycin-resistant mTOR activity is required for sensory axon regeneration induced by a conditioning lesion. *eNeuro* **3**, ENEURO.0358-16.2016 (2017). doi: 10.1523/ENEURO.0358-16.2016; pmid: 28101526
18. T. C. Jackson, A. Rani, A. Kumar, T. C. Foster, Regional hippocampal differences in AKT survival signaling across the lifespan: Implications for CA1 vulnerability with aging. *Cell Death Differ.* **16**, 439–448 (2009). doi: 10.1038/cdd.2008.171; pmid: 19039330
19. B. Du *et al*., CD4$^+$ αβ T cell infiltration into the leptomeninges of lumbar dorsal roots contributes to the transition from acute to chronic mechanical allodynia after adult rat tibial nerve injuries. *J. Neuroinflammation* **15**, 81 (2018). doi: 10.1186/s12974-018-1115-7; pmid: 29544518
20. J. G. Cyster *et al*., Follicular stromal cells and lymphocyte homing to follicles. *Immunol. Rev.* **176**, 181–193 (2003). doi: 10.1034/j.1600-065X.2000.00618.x; pmid: 11043777
21. B.-C. Jiang *et al*., CXCL13 drives spinal astrocyte activation and neuropathic pain via CXCR5. *J. Clin. Invest.* **126**, 745–761 (2016). doi: 10.1172/JCI81950; pmid: 26752644
22. X.-B. Wu *et al*., CXCL13/CXCR5 enhances sodium channel Nav1.8 current density via p38 MAP kinase in primary sensory neurons following inflammatory pain. *Sci. Rep.* **6**, 34836 (2016). doi: 10.1038/srep34836; pmid: 27708397
23. R. He *et al*., Follicular CXCR5- expressing CD8(+) T cells curtail chronic viral infection. *Nature* **537**, 412–428 (2016). doi: 10.1038/nature19317; pmid: 27501245
24. D. Yu, L. Ye, A portrait of CXCR5+ follicular cytotoxic CD8$^+$ T cells. *Trends Immunol.* **39**, 965–979 (2018). doi: 10.1016/j.it.2018.10.002; pmid: 30377045
25. B. W. Dulken *et al*., Single-cell analysis reveals T cell infiltration in old neurogenic niches. *Nature* **571**, 205–210 (2019). doi: 10.1038/s41586-019-1362-5; pmid: 31270459
26. B. Coder *et al*., Friend or foe: The dichotomous impact of T cells on neuro-de/re-generation during aging. *Oncotarget* **8**, 7116–7137 (2017). doi: 10.18632/oncotarget.12572; pmid: 27738345
27. G. Nardo *et al*., Counteracting roles of MHCI and CD8$^+$ T cells in the peripheral and central nervous system of ALS SOD1^{G93A} mice. *Mol. Neurodegener.* **13**, 42 (2018). doi: 10.1186/s13024-018-0271-7; pmid: 30092791
28. M. Daglas *et al*., Activated CD8$^+$ T Cells Cause Long-Term Neurological Impairment after Traumatic Brain Injury in Mice. *Cell Rep.* **29**, 1178–1191.e6 (2019). doi: 10.1016/j.celrep.2019.09.046; pmid: 31665632
29. A. L. Oliveira *et al*., A role for MHC class I molecules in synaptic plasticity and regeneration of neurons after axotomy. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17843–17848 (2004). doi: 10.1073/pnas.0408154101; pmid: 15591351
30. M. S. Joseph *et al*., Transgenic mice with enhanced neuronal major histocompatibility complex class I expression recover locomotor function better after spinal cord injury. *J. Neurosci. Res.* **89**, 365–372 (2011). doi: 10.1002/jnr.22557; pmid: 21259323
31. J. E. Knickelbein *et al*., Noncytotoxic lytic granule-mediated CD8$^+$ T cell inhibition of HSV-1 reactivation from neuronal latency. *Science* **322**, 268–271 (2008). doi: 10.1126/science.1164164; pmid: 18845757
32. S. G. Meuth *et al*., Cytotoxic CD8$^+$ T cell-neuron interactions: Perforin-dependent electrical silencing precedes but is not causally linked to neuronal cell death. *J. Neurosci.* **29**, 15397–15409 (2009). doi: 10.1523/JNEUROSCI.4339-09.2009; pmid: 20007464
33. M. D'Amelio, V. Cavallucci, F. Cecconi, Neuronal caspase-3 signaling: Not only cell death. *Cell Death Differ.* **17**, 1104–1114 (2010). doi: 10.1038/cdd.2009.180; pmid: 19960023
34. H. A. Wols *et al*., Migration of immature and mature B cells in the aged microenvironment. *Immunology* **129**, 278–290 (2010). doi: 10.1111/j.1365-2567.2009.03182.x; pmid: 19845796
35. J. S. Lefebvre *et al*., The aged microenvironment contributes to the age-related functional defects of CD4 T cells in mice. *Aging Cell* **11**, 732–740 (2012). doi: 10.1111/j.1474-9726.2012.00836.x; pmid: 22607653
36. E. Klimatcheva *et al*., CXCL13 antibody for the treatment of autoimmune disorders. *BMC Immunol.* **16**, 6 (2015). doi: 10.1186/s12865-015-0068-1; pmid: 25879435
37. A. R. Bauder, T. A. Ferguson, Reproducible mouse sciatic nerve crush and subsequent assessment of regeneration by whole mount muscle analysis. *J. Vis. Exp.*3606 (2012). doi: 10.3791/3606; pmid: 22395197
38. T. H. Hutson *et al*., Cbp-dependent histone acetylation mediates axon regeneration induced by environmental enrichment in rodent spinal cord injury models. *Sci. Transl. Med.* **11**, eaaw2064 (2019). doi: 10.1126/scitranslmed.aaw2064; pmid: 30971452
39. A. Hervera *et al*., Reactive oxygen species regulate axonal regeneration through the release of exosomal NADPH oxidase 2 complexes into injured axons. *Nat. Cell Biol.* **20**, 307–319 (2018). doi: 10.1038/s41556-018-0039-x; pmid: 29434374
40. Z. Keren *et al*., B-cell depletion reactivates B lymphopoiesis in the BM and rejuvenates the B lineage in aging. *Blood* **117**, 3104–3112 (2011). doi: 10.1182/blood-2010-09-307983; pmid: 21228330
41. B. Chen *et al*., Reactivation of dormant relay pathways in injured spinal cord by KCC2 manipulations. *Cell* **174**, 521–535.e13 (2018). doi: 10.1016/j.cell.2018.06.005; pmid: 30033363
42. S. J. Gladman *et al*., Improved outcome after peripheral nerve injury in mice with increased levels of endogenous ω-3 polyunsaturated fatty acids. *J. Neurosci.* **32**, 563–571 (2012). doi: 10.1523/JNEUROSCI.3371-11.2012; pmid: 22238091
43. A. C. Ogbonna, A. K. Clark, M. Malcangio, Development of monosodium acetate-induced osteoarthritis and inflammatory pain in ageing mice. *Age* **37**, 54 (2015). doi: 10.1007/s11357-015-9792-y; pmid: 25971876
44. N. Agarwal *et al*., Evoked hypoalgesia is accompanied by tonic pain and immune cell infiltration in the dorsal root ganglia at late stages of diabetic neuropathy in mice. *Mol. Pain* **14**, 1744806918817975 (2018). doi: 10.1177/1744806918817975; pmid: 30453826
45. S. M. Fleming *et al*., Early and progressive sensorimotor anomalies in mice overexpressing wild-type human α-synuclein. *J. Neurosci.* **24**, 9434–9440 (2004). doi: 10.1523/JNEUROSCI.3080-04.2004; pmid: 15496679

ACKNOWLEDGMENTS

We thank M. Jucker for providing several aged wild-type mice, M. Merkenschlager and J. Fawcett for their critical feedback on the manuscript, T. Arnon and her team for providing material from *Cxcr5*$^{-/-}$ mice, the LMS/NIHR Imperial Biomedical Research Center Flow Cytometry Facility for support, and N. Haberman for advice on statistical analysis. **Funding:** Supported by Wings for Life, The Rosetrees Trust, the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre, Guarantors of Brain (L.Z.), and start-up funds from the Department of Brain Sciences, BRC funds, Imperial College London (S.D.G.) and Wings for Life grant WFL-NL-25/20 (J.V.). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the UK Department of Health. **Author contributions:** L.Z. designed and performed experiments and data analysis and wrote the paper; G.K. designed and performed experiments and performed data analysis; I.P. designed experiments and performed data analysis; M.D. performed data analysis; G.C. performed experiments; F.D.V. performed experiments; M.T.C. performed data analysis; J.S.C. performed experiments; Z.Y. performed experiments; V.L. provided experimental advice and edited the paper; J.B. provided experimental advice; R.P. provided experimental advice and edited the paper; J.V. provided research material for experiments and edited the paper; F.D.W. provided research material for experiments; C.P. provided research material for experiments and experimental advice; C.L.C. provided research material and experimental advice; J.S. provided experimental advice and edited the paper; M.B. provided experimental advice and edited the paper; S.D.G. designed experiments, provided funding, and wrote the paper. **Competing interests:** L.Z. and S.D.G. have a pending patent application related to this work titled "Use of CXCL13 Binding Molecules to Promote Peripheral Nerve Regeneration." The authors declare that they have no other competing interests. **Data and materials availability:** The raw RNA-seq data used and generated in this study have been deposited in the GEO database under accession numbers GSE138769 and GSE149770. All other data are available in the main text or supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abd5926
Figs. S1 to S19
Table S1
Data S1 to S3

Submitted 3 July 2020; resubmitted 22 September 2021
Accepted 9 March 2022
10.1126/science.abd5926

RESEARCH ARTICLES

QUANTUM SIMULATION

Quantum gas microscopy of Kardar-Parisi-Zhang superdiffusion

David Wei[1,2], Antonio Rubio-Abadal[1,2]†, Bingtian Ye[3], Francisco Machado[3,4], Jack Kemp[3], Kritsana Srakaew[1,2], Simon Hollerith[1,2], Jun Rui[1,2]‡, Sarang Gopalakrishnan[5,6], Norman Y. Yao[3,4], Immanuel Bloch[1,2,7], Johannes Zeiher[1,2]*

The Kardar-Parisi-Zhang (KPZ) universality class describes the coarse-grained behavior of a wealth of classical stochastic models. Surprisingly, KPZ universality was recently conjectured to also describe spin transport in the one-dimensional quantum Heisenberg model. We tested this conjecture by experimentally probing transport in a cold-atom quantum simulator via the relaxation of domain walls in spin chains of up to 50 spins. We found that domain-wall relaxation is indeed governed by the KPZ dynamical exponent $z = 3/2$ and that the occurrence of KPZ scaling requires both integrability and a nonabelian SU(2) symmetry. Finally, we leveraged the single-spin–sensitive detection enabled by the quantum gas microscope to measure an observable based on spin-transport statistics. Our results yield a clear signature of the nonlinearity that is a hallmark of KPZ universality.

In many-particle systems, hydrodynamics (*1*, *2*) naturally emerges upon coarse graining from the microscopic equations of motion in both classical and quantum systems (*3*–*7*). A celebrated example of the emergence of hydrodynamics is the so-called Kardar-Parisi-Zhang (KPZ) equation, which governs an abundance of disparate phenomena ranging from interface growth to the shapes of polymers, the spreading of sound waves in fluids, and the growth of entanglement in random quantum circuits (*8*–*11*). A single equation can describe so many distinct physical systems because of the notion of universality (*12*). Indeed, the physical properties of systems described by the KPZ equation follow scaling functions that were exactly computed in a milestone mathematical achievement (*13*). All of these canonical KPZ systems share the feature that their dynamics is highly robust: KPZ scaling requires no particular symmetries and occurs in the presence of external noise (*14*).

Recent efforts have focused on the prediction that KPZ hydrodynamics should emerge in an entirely distinct setting: the spin-½ quantum Heisenberg chain at infinite temperature (*15*–*21*). The appearance of KPZ scaling in such a quantum spin chain is a surprise because the chains lack the properties shared by all canonical KPZ systems. In particular, noise is absent and the KPZ scaling is fragile—it requires integrability (*6*, *22*–*24*), a feature of the Heisenberg spin chain that does not apply to canonical examples of KPZ. The very appearance of nontrivial dynamical scaling in high-temperature near-equilibrium states is unexpected: The theory of dynamical critical phenomena (*25*) does not apply because high-temperature states are far from any equilibrium critical point. Moreover, nonlinear hydrodynamics predicts that transport in spin chains at nonzero temperature should be diffusive (*26*).

Notably, recent conjectures suggest that the emergence of KPZ hydrodynamics in quantum many-body systems could be substantially more general and may apply to any integrable model that exhibits nonabelian symmetry (*27*, *28*), suggesting a fundamental difference from the mechanisms of canonical KPZ (*21*, *27*–*31*). To date, a full theory of KPZ hydrodynamics in the Heisenberg model remains elusive (*26*, *32*). In light of these questions, experimental characterization of the anomalous dynamical exponents of spin transport has been the subject of widespread effort (*33*–*36*). However, because a superdiffusive exponent can arise from a number of distinct microscopic origins (*37*), it is essential to characterize a system beyond the dynamical exponent to establish its hydrodynamical universality class.

In this work, we explore the superdiffusive dynamics of the ferromagnetic Heisenberg model by using a quantum gas microscope with single-site resolution and single-spin–sensitive detection in spin chains of up to 50 spins. Our main results are threefold. First, we observe superdiffusive spin transport with the dynamical exponent $z = 3/2$, consistent with KPZ hydrodynamics. Second, we demonstrate that both integrability and nonabelian symmetry are essential for observing superdiffusion: Breaking integrability by tuning dimensionality restores diffusion, and breaking the symmetry by preparing an initial state with net magnetization leads to ballistic transport (Fig. 1A). Finally, leveraging the ability of our experimental setup to detect spin-resolved snapshots of the entire sample, we map the shot-to-shot dynamical fluctuations (i.e., the "full counting statistics") of the magnetization. These fluctuations carry clear signatures of the intrinsic nonlinearity associated with KPZ hydrodynamics (*38*) and distinguish it from other potential mechanisms for superdiffusion, such as Lévy flights (*26*).

Experimental system

In our experiment, we probed the transport dynamics of bosonic ^{87}Rb atoms trapped in an optical lattice; the atoms occupy the two hyperfine ground states $|\uparrow\rangle = |F = 1, m_F = -1\rangle$ and $|\downarrow\rangle = |F = 2, m_F = -2\rangle$ (where F is the total angular momentum and m_F is the Zeeman sublevel), and their dynamics are captured by a two-species Bose-Hubbard model with on-site interaction U and tunnel coupling $\tilde{t}$. At unit filling and in the limit of strong interactions, the direct tunneling between lattice sites is suppressed, and spin dynamics occur via second-order spin-exchange. The system can be mapped to the spin-½ XXZ model for $|\uparrow\rangle$ and $|\downarrow\rangle$ (*39*, *40*), and, in one dimension, is described by the Hamiltonian

$$\hat{H} = -J\sum_j \left(\hat{S}_j^x\hat{S}_{j+1}^x + \hat{S}_j^y\hat{S}_{j+1}^y + \Delta\hat{S}_j^z\hat{S}_{j+1}^z\right) \quad (1)$$

where Δ quantifies the interaction anisotropy, $J = 4\tilde{t}^2/U$ characterizes the spin-exchange coupling, and j denotes the lattice sites. In our system, the atomic scattering properties yield $\Delta \approx 1$, and the system maps to the isotropic ferromagnetic Heisenberg model (*41*).

We began our experiment by loading a spin-polarized two-dimensional (2D) degenerate gas of ~2000 atoms into a square optical lattice with a spacing of $a = 532$ nm. We achieved a homogeneous box potential over an area of 50 sites by 22 sites by additionally projecting light at a wavelength of 670 nm with a digital micromirror device (DMD), preparing a Mott insulator with a filling of $n_0 = 0.93(1)$ in this box [see details in (*41*)]. Local spin control was realized using light at a wavelength of 787 nm on the DMD (*42*) to apply a site-resolved differential light shift between $|\uparrow\rangle$ and $|\downarrow\rangle$; subsequent microwave driving allowed for local flips of the spatially addressed spins.

[1]Max-Planck-Institut für Quantenoptik, 85748 Garching, Germany. [2]Munich Center for Quantum Science and Technology (MCQST), 80799 Munich, Germany. [3]Department of Physics, University of California, Berkeley, CA 94720, USA. [4]Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [5]Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA. [6]Department of Physics and Astronomy, College of Staten Island, Staten Island, NY 10314, USA. [7]Fakultät für Physik, Ludwig-Maximilians-Universität, 80799 Munich, Germany.
*Corresponding author. Email: johannes.zeiher@mpq.mpg.de
†Present address: Institut de Ciències Fotòniques (ICFO), The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain.
‡Present address: Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, Anhui 230026, China.

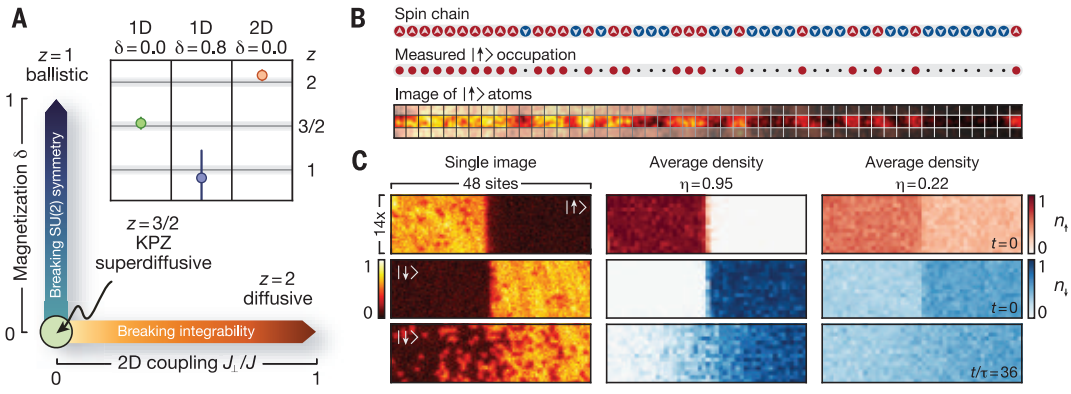


Fig. 1. Hydrodynamic transport in Heisenberg chains and schematic of the experimental system. (**A**) Dynamical exponents for finite-temperature Heisenberg chains. Whereas integrable systems typically display ballistic transport (magnetized chains, $\delta > 0$), nonintegrable systems are generically diffusive (2D Heisenberg model, $J_\perp > 0$). For unmagnetized Heisenberg chains, transport is expected to fall into the KPZ universality class with a superdiffusive exponent $z = 3/2$. (Inset) By measuring polarization transfer $P(t)$ across a domain wall, we directly observe these transport regimes: superdiffusion in the unmagnetized case (green), ballistic transport at finite net magnetization (blue), and diffusion in two dimensions (orange). Exponents are extracted by fitting $P(t) \propto t^{1/z}$; for the ballistic case, we additionally fit a vertical intercept to account for transient initial-time dynamics. Error bars denote SD of the fit. (**B**) In each experimental run, we measure the spin states of a Heisenberg chain (top) by removing one spin species (center) and imaging the atomic site occupation (bottom). (**C**) The Heisenberg chains are achieved in a 2D atomic Mott insulator (analysis region depicted) with controllable interchain coupling. Our setup allows us to prepare domain walls with high-purity η (left and middle columns) and low-purity η (right). We measure the time evolution of both $|\uparrow\rangle$ (top) and $|\downarrow\rangle$ (middle and bottom rows) atoms to extract the polarization transfer.

Such quantum control enabled us to prepare spin domain walls (*16*, *17*, *43*, *44*) by spatially addressing half of the system. Subsequently, we prepared high-entropy states by globally rotating the spins away from the S^z axis by using a resonant microwave pulse and then locally dephasing them by projecting a site-to-site random spin-dependent potential, which we modified from shot to shot (*41*) (Fig. 1C). More precisely, our experiments focused on tracking spin dynamics, starting from a class of initial states containing a spin domain wall with magnetization difference 2η in the middle of the spin chain—i.e., half of the system has magnetization η; the other half has magnetization $-\eta$. In the infinite-temperature limit, $\eta \to 0$, the relaxation of such states yields linear response transport coefficients, as the derivative of the spin profile is precisely the dynamical spin structure factor (*16*, *17*).

To probe 1D spin dynamics in our system, we rapidly quenched the lattice depth along 1D tubes with a length of 50 sites, which suddenly increased the spin-exchange coupling from zero to $J/\hbar = 64(1)\mathrm{s}^{-1}$ (where $\hbar$ is Planck's constant divided by 2π). After tracking the spin dynamics for up to ~45 spin-exchange times $\tau = \hbar/J$, we removed one spin component and measured the remaining occupation via fluorescence imaging (Fig. 1B).

Superdiffusive spin transport

To explore the nature of anomalous spin transport in the 1D Heisenberg model, we initialize the spins in a high-entropy domain-wall state with $\eta = 0.22(2)$. We characterize the subsequent spin transport by measuring the polarization transfer, $P(t)$, defined as the average total number of spins that have crossed the domain wall by time t (*41*). The emergence of hydrodynamics is characterized by the power-law scaling of $P(t) \sim t^{1/z}$ and immediately enables us to extract the underlying dynamical exponent z. As depicted in Fig. 2A, the data exhibit a superdiffusive exponent, $z = 1.54(7)$, consistent with KPZ scaling. By comparison, neither a diffusive ($z = 2$) nor a ballistic ($z = 1$) exponent accurately captures the observed dynamics (Fig. 2B) (*41*). Somewhat surprisingly, we also observe a superdiffusive exponent of $z = 1.45(5)$ upon changing the initial state to a near-pure domain wall with $\eta = 0.95(2)$ (fig. S8) (*26*, *41*, *44*–*46*).

To further explore the superdiffusive dynamics, we investigate the spatially resolved spin profiles at $\eta = 0.22(2)$. Our experimental observations are in quantitative agreement with simulations based on tensor-network numerical techniques (*41*, *46*, *47*) and conform to KPZ dynamics (Fig. 2A). Crucially, when appropriately rescaled by the dynamical exponent, all of the observed spatiotemporal profiles collapse onto a scaling form consistent with the KPZ scaling function (Fig. 2C).

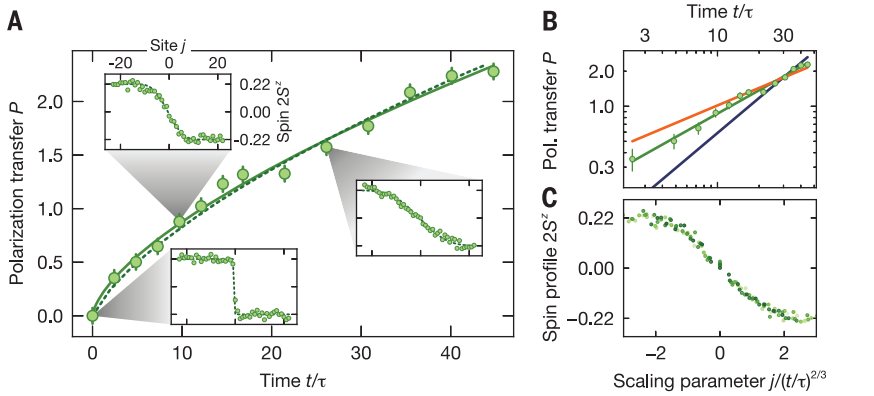


Fig. 2. Superdiffusive spin transport in a high-temperature Heisenberg chain. (**A**) The polarization transfer for a domain-wall initial state with a contrast of $\eta = 0.22(2)$ grows as a power law [$P(t) \propto t^{1/z}$] with a fitted exponent $z = 1.54(7)$ (solid line), indicating superdiffusive transport. The experimental data are consistent with numerical Heisenberg-model simulations (*41*) (dashed line). The insets show the averaged spin profiles $2S_j^z(t)$ at times t/τ = 0, 10, 26, which are compared to simulations (dashed lines). (**B**) Polarization transfer in a double-logarithmic plot. The solid lines are power-law fits with fixed exponents, where a distinction between z = 3/2 (green) and both z = 2 (brown) and z = 1 (blue) is visible. (**C**) When rescaling time by the inverse dynamical exponent, the spatial spin profiles at times t/τ = 5 to 35 (light to dark green) collapse to a characteristic shape consistent with the integrated KPZ function. Error bars denote SEM.

Microscopic origins of superdiffusion

To understand why the combination of integrability and nonabelian symmetry leads to emergent superdiffusive transport, it is instructive to first consider the transport dynamics on

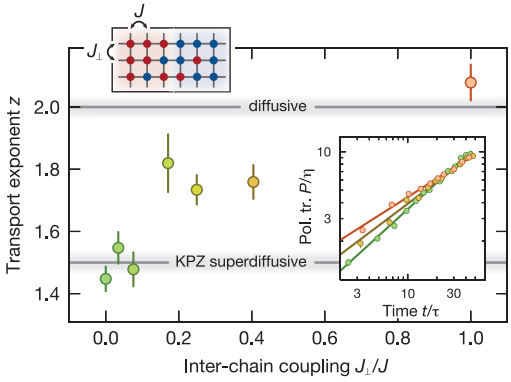


Fig. 3. Evolution toward diffusive transport under a breakdown of integrability. Fitted power-law exponent z for the spin polarization transfer at different coupling strengths between individual 1D chains with initial domain walls with $\eta \approx 0.9$ (*41*). Starting from superdiffusive transport in the purely 1D case, $z = 1.45(5)$, increased interchain coupling breaks the integrability of the system and leads to a crossover toward diffusive transport, reaching $z = 2.08(4)$ in the 2D case, as generically expected for nonintegrable systems. The inset depicts the normalized polarization transfer $P(t)/\eta$ for $J_\perp/J = 0, 0.40(1)$ and $1.00(5)$ (green to orange). Error bars denote SD of the fit.

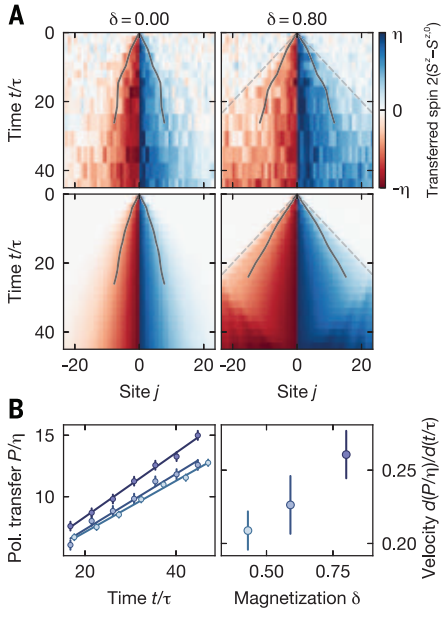


Fig. 4. Ballistic spin dynamics under broken SU(2) symmetry. (**A**) Averaged experimental (top) and numerical (bottom) spin profiles $S_j^z(t)$, from which the initial profile $S_j^{z,0}$ is subtracted. (Left) Unmagnetized low-purity domain wall, $\delta = 0, \eta = 0.22(2)$ (from Fig. 2). Spin transport results from increased spin profile width, which scales with the superdiffusive dynamical exponent. The solid lines indicate the position j where the spin profile crosses $\left|2S_j^z(t) - \delta\right|/\eta = 0.4$; because the profiles themselves are scale invariant, the position of any S^z value follows the z = 3/2 scaling. (Right) Magnetized domain wall, $\delta = 0.80(1), \eta = 0.12(1)$. At the outer edge, the contribution of magnons is visible, transporting spin with the speed of the spin "light cone" (dashed line), which was measured with a quantum walk (*41*). Most of the spin is carried by quasiparticles within the light cone, and thus the width of the profile (solid line) grows faster than in the unmagnetized case. The numerical simulation shows a qualitatively similar behavior. At $t/\tau \simeq 25$ the magnons reach the system edge and are reflected. (**B**) To extract the ballistic polarization-transfer velocity, we linearly fit the normalized polarization transfer after a crossover time, $t/\tau > 16$ (left). We observe transfer velocity growth as the initial domain-wall magnetization δ increases (right, light to dark blue). Error bars denote SD of the fit.

top of a small net magnetization background (*18*–*20*, *48*). In our experiments, this corresponds to preparing domain walls with a finite overall magnetization δ—i.e., half of the system has a magnetization η + δ; the other half has −η + δ. Stable quasiparticles then render spin transport ballistic (Fig. 1A), leading to a characteristic polarization-transfer rate that scales linearly with net magnetization δ (*20*). Even when δ = 0 on average, random local fluctuations of the magnetization will be present; thus, the net magnetization in a typical region of size ℓ will scale as $1/\sqrt{\ell}$. Therefore, the average spin-transport rate across a region of size ℓ also scales as $1/\sqrt{\ell}$, implying that the transport time across the region scales as $\ell/(1/\sqrt{\ell}) \sim \ell^{3/2}$, precisely yielding the KPZ exponent z = 3/2 [see details in (*41*)].

This intuitive analysis suggests two key requirements for superdiffusive transport: (i) integrability ensures the presence of stable quasiparticles that move ballistically, and (ii) the presence of nonabelian SU(2) symmetry makes the characteristic velocity of the ballistic contribution to spin transport vanish. We can experimentally probe these requirements by individually breaking either the integrability or the SU(2) symmetry of the system.

To break integrability, we turn on a finite interchain coupling $J_\perp$ by lowering the lattice depth orthogonal to the 1D spin chains, which effectively causes the system to become 2D (*49*, *50*). We measure the dependence of the polarization transfer on the interchain coupling, starting from an unmagnetized domain wall [$\eta \approx 0.9$ (*41*), δ = 0]. As shown in Fig. 3, the extracted dynamical exponents exhibit a clear flow from superdiffusive transport when $J_\perp = 0$ to diffusive transport, $z = 2.08(4)$, when $J_\perp = J$. Notably, for $J_\perp/J \lesssim 0.1$ integrability is strictly broken, but the transport dynamics remain consistent with superdiffusion within experimentally accessible time scales. This observation bolsters recent theoretical expectations, which suggest that superdiffusion can be particularly robust to perturbations that do not break the nonabelian symmetry (*51*).

Next, we explore the effect of breaking the underlying SU(2) symmetry by using initial states with finite net magnetization δ (*41*). Working with an imbalanced domain-wall initial state [$\eta = 0.12(1), \delta = 0.80(1)$], we observe two main differences from the unmagnetized (δ = 0) case (Fig. 4A). First, the polarization profile exhibits a fast ballistic component that follows the spin "light cone" of the dynamics (with a speed of 1/τ; dashed line in Fig. 4A). This contribution arises from the fastest quasiparticles, which transport spin above the magnetized background (*41*, *52*). Second, within this light cone, polarization spreads substantially faster than in the unmagnetized case; this comprises the bulk of the spin transport and is mediated by slower-moving, net-magnetization–carrying quasiparticles.

At early times, the polarization-transfer dynamics exhibit a superdiffusive power law before crossing over to linear ballistic transport at later times (*41*). In particular, by fitting a power law to the late-time data, $t/\tau > 16$, we extract a dynamical exponent $z = 0.9(3)$, consistent with ballistic spin transport (Fig. 4B). Although our results agree qualitatively with numerical simulations of the Heisenberg model, the magnitude of the measured polarization transfer is smaller; this can be understood as resulting from the presence of hole defects in the initial state (*41*, *53*). In addition to verifying the ballistic nature of the spin dynamics, we can also directly extract the velocity of the underlying quasiparticles. By controlling the overall magnetization of the initial state, we observe the expected increase of the velocity with δ (Fig. 4B), an essential component for understanding the presence of KPZ superdiffusion in spin chains (*20*).

Observing KPZ hydrodynamics

Our previous observations have focused on characterizing superdiffusive spin transport. However, from the perspective of observing KPZ universality, this is insufficient, as multiple different classes of hydrodynamics can exhibit the same dynamical exponent of z = 3/2. To distinguish these classes, we go beyond measurements of the average polarization transfer and analyze the full distribution function of the polarization transfer $\Pr(P; t)$ across snapshots (*38*). This distribution function can distinguish KPZ from potential alternatives

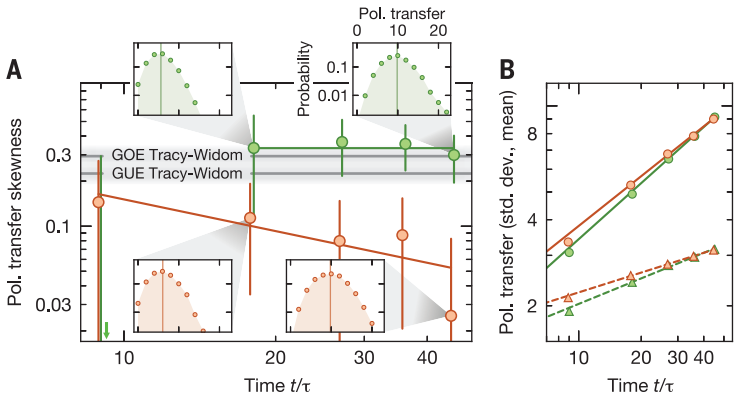


Fig. 5. Distribution function of polarization transfer. (**A**) The probability distribution asymmetry of the polarization transfer expected for KPZ transport is quantified by the skewness. We compare the pure domain-wall dynamics in the 1D case (green) with the nonintegrable 2D case at $J_\perp/J = 0.25(1)$ (orange). Whereas the 2D case becomes symmetric at late times, the 1D distribution remains asymmetric with a skewness of 0.33(8). Gray lines indicate the skewness of the Gaussian-orthogonal-ensemble (GOE) and Gaussian-unitary-ensemble (GUE) Tracy-Widom (TW) distributions (*41*, *54*). Colored lines serve as guides to the eye. (Insets) Probability distributions of the polarization transfer on a logarithmic scale. The vertical line marks the mean of the distribution. (**B**) The mean (circles) of the polarization transfer is consistent with the data shown in Fig. 3 and scales with the power-law (solid lines) exponent [$1/z = 0.67(1)$ in one dimension; $1/z = 0.60(2)$ in two dimensions]. The standard deviation (triangles) features another characteristic transport exponent [the growth exponent (*55*)], which is consistent with the extracted power-law (dashed lines) exponent [$\beta = 0.31(1)$ in one dimension; $\beta = 0.24(1)$ in two dimensions]. Error bars denote the SD obtained from a bootstrap analysis.

such as Lévy flights: For all linear processes (such as Lévy flights or time-rescaled diffusion), the fluctuations of P at late times are necessarily symmetric about the mean; for KPZ, the limiting distribution $\Pr(P; t\to\infty)$ is the Tracy-Widom (TW) distribution (*41*), which is strongly asymmetric (*10*, *54*).

Measuring the statistics of the polarization-transfer distribution therefore gives us a direct experimental observable to discern the underlying hydrodynamical transport equations. This analysis fundamentally relies on the single-shot nature and single-spin sensitivity of our quantum gas microscope. As we measure the occupation of a single-spin species per snapshot, we approximate the polarization-transfer statistics by the statistics for the single-species atom-number transfer, $N_T^{\uparrow(\downarrow)} \approx P/2$, where $N_T^{\uparrow}$ is the number of $|\uparrow\rangle$ atoms on the side of the domain wall initialized with the opposite spin $|\downarrow\rangle$. We quantify the asymmetry of the distribution about its mean $\bar{P}$ by its skewness $(\mu_3(t) - \mu_3(0))/(\mu_2(t) - \mu_2(0))^{3/2}$ (*41*), where $\mu_k(t) = \sum_P (P-\bar{P})^k \Pr(P;t)$ denotes the kth central moment of the distribution.

To begin, we characterize the skewness of the polarization transfer starting from a high-purity domain wall [$\eta = 0.89(1)$, $\delta = 0$] for a 2D geometry with an interchain coupling strength $J_\perp/J = 0.25(1)$. The skewness of the polarization transfer distribution is small overall and is most consistent with a decay toward zero (Fig. 5) (*41*), which corresponds to a fully symmetric distribution and is expected for linear diffusive processes exhibited by the nonintegrable 2D Heisenberg model.

If the 1D Heisenberg model is actually governed by nonlinear KPZ hydrodynamics, one would expect markedly distinctive behavior for the skewness as a function of time. In particular, the nonlinearity of the KPZ equation would lead to a finite skewness, which is constant over time. We indeed observe that the skewness saturates to a finite value of 0.33(8) when starting from an initial state with $\eta = 0.91(2)$ and $\delta = 0$ (Fig. 5). In agreement with numerical simulations, this value is consistent with the skewness of the Gaussian-orthogonal-ensemble (GOE) TW distribution, 0.294 (*54*). Our experiment directly rules out linear transport processes and thus provides a strong indication that transport in the 1D quantum Heisenberg chain is governed by KPZ hydrodynamics.

Discussion and conclusion

Our results support the theoretical conjecture that spin transport in the 1D Heisenberg model belongs to the KPZ universality class, with a superdiffusive transport exponent $z = 3/2$. We have experimentally demonstrated that both integrability and nonabelian symmetry are essential for stabilizing superdiffusive transport. Moreover, we exploit the single-spin sensitivity of our setup to extract the full distribution function of the polarization transfer. This distribution function exhibits a large skewness that does not decay in time, demonstrating that spin transport in this system belongs to a strongly coupled, nonlinear dynamical universality class.

Our work builds and expands on recent experimental explorations of Heisenberg-model spin dynamics. These experiments include neutron scattering studies of the quantum material $KCuF_3$ (*36*), as well as experiments that probe the relaxation of spin-spiral initial states in ultracold gases (*33*–*35*). In the 1D Heisenberg model, the relaxation of such spin-spiral states is nongeneric because they are approximate eigenstates in the long-wavelength limit (*26*). Empirically, spin spirals relax with a diffusive exponent $z = 2$. By considering a more generic family of domain-wall initial states, we are able to directly probe (and controllably move away from) the high-temperature linear-response limit where KPZ transport is thought to occur.

Our results open the door to a number of future directions. First, the discrepancy between the relaxation of domain walls and spin spirals (away from linear response) indicates that relaxation in integrable systems is generally strongly dependent on initial state, although we currently lack a theory of this nonlinear regime. Second, the robustness of our results along the crossover from the Heisenberg to the (nonintegrable) Bose-Hubbard regime remains to be fully understood (*51*). In this context, a comparison between the nonintegrable Bose-Hubbard model and the integrable Fermi-Hubbard model (*53*) could be of particular interest. Finally, the observable we introduced to capture fluctuation effects—namely, the statistics of single shots of the polarization transfer—promises to be a powerful diagnostic tool for new phases of interacting quantum systems. A theory of this quantity already exists for the KPZ universality class, but developing a more general theory of such transport fluctuations will be an important task for future theoretical work.

Note added in proof: During the completion of this manuscript, we became aware of related work to observe superdiffusive transport in a long-range interacting ion chain (*37*).

REFERENCES AND NOTES

1. H. Spohn, *Large Scale Dynamics of Interacting Particles* (Springer, 2012).
2. G. Birkhoff, *Hydrodynamics* (Princeton Univ. Press, 2015).
3. R. E. Wyatt, *Quantum Dynamics with Trajectories: Introduction to Quantum Hydrodynamics*, vol. 28 (Springer, 2006).
4. S. Mukerjee, V. Oganesyan, D. Huse, *Phys. Rev. B* **73**, 035113 (2006).
5. L. Erdős, B. Schlein, H.-T. Yau, *Phys. Rev. Lett.* **98**, 040404 (2007).
6. O. A. Castro-Alvaredo, B. Doyon, T. Yoshimura, *Phys. Rev. X* **6**, 041065 (2016).
7. C. Zu *et al.*, *Nature* **597**, 45–50 (2021).
8. M. Kardar, G. Parisi, Y.-C. Zhang, *Phys. Rev. Lett.* **56**, 889–892 (1986).
9. T. Halpin-Healy, K. Takeuchi, *J. Stat. Phys.* **160**, 794–814 (2015).
10. H. Spohn, *J. Stat. Mech.* **2020**, 044001 (2020).
11. A. Nahum, J. Ruhman, S. Vijay, J. Haah, *Phys. Rev. X* **7**, 031016 (2017).
12. L. P. Kadanoff, *Physica A* **163**, 1–14 (1990).
13. M. Prähofer, H. Spohn, *J. Stat. Phys.* **115**, 255–279 (2004).

14. M. Gubinelli, N. Perkowski, *Commun. Math. Phys.* **349**, 165–269 (2017).
15. M. Znidarič, *Phys. Rev. Lett.* **106**, 220601 (2011).
16. M. Ljubotina, M. Žnidarič, T. Prosen, *Nat. Commun.* **8**, 16117 (2017).
17. M. Ljubotina, M. Žnidarič, T. Prosen, *Phys. Rev. Lett.* **122**, 210602 (2019).
18. S. Gopalakrishnan, R. Vasseur, *Phys. Rev. Lett.* **122**, 127202 (2019).
19. J. De Nardis, M. Medenjak, C. Karrasch, E. Ilievski, *Phys. Rev. Lett.* **123**, 186601 (2019).
20. S. Gopalakrishnan, R. Vasseur, B. Ware, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16250–16255 (2019).
21. V. B. Bulchandani, *Phys. Rev. B* **101**, 041411 (2020).
22. B. Bertini, M. Collura, J. De Nardis, M. Fagotti, *Phys. Rev. Lett.* **117**, 207201 (2016).
23. E. Ilievski, J. De Nardis, *Phys. Rev. Lett.* **119**, 020602 (2017).
24. V. B. Bulchandani, R. Vasseur, C. Karrasch, J. E. Moore, *Phys. Rev. B* **97**, 045407 (2018).
25. P. C. Hohenberg, B. I. Halperin, *Rev. Mod. Phys.* **49**, 435–479 (1977).
26. V. B. Bulchandani, S. Gopalakrishnan, E. Ilievski, *J. Stat. Mech.* **2021**, 084001 (2021).
27. Ž. Krajnik, T. Prosen, *J. Stat. Phys.* **179**, 110–130 (2020).
28. E. Ilievski, J. De Nardis, S. Gopalakrishnan, R. Vasseur, B. Ware, *Phys. Rev. X* **11**, 031023 (2021).
29. T. Prosen, B. Zunkovič, *Phys. Rev. Lett.* **111**, 040602 (2013).
30. A. Das, M. Kulkarni, H. Spohn, A. Dhar, *Phys. Rev. E* **100**, 042116 (2019).
31. Ž. Krajnik, E. Ilievski, T. Prosen, *SciPost Phys.* **9**, 038 (2020).
32. B. Bertini *et al.*, *Rev. Mod. Phys.* **93**, 025003 (2021).
33. S. Hild *et al.*, *Phys. Rev. Lett.* **113**, 147205 (2014).
34. P. N. Jepsen *et al.*, *Nature* **588**, 403–407 (2020).
35. P. N. Jepsen *et al.*, *Phys. Rev. X* **11**, 041054 (2021).
36. A. Scheie *et al.*, *Nat. Phys.* **17**, 726–730 (2021).
37. M. K. Joshi *et al.*, *Science* **376**, 720–724 (2022).
38. A. K. Hartmann, P. Le Doussal, S. N. Majumdar, A. Rosso, G. Schehr, *Europhys. Lett.* **121**, 67004 (2018).
39. L.-M. Duan, E. Demler, M. D. Lukin, *Phys. Rev. Lett.* **91**, 090402 (2003).
40. A. B. Kuklov, B. V. Svistunov, *Phys. Rev. Lett.* **90**, 100401 (2003).
41. See supplementary materials.
42. T. Fukuhara *et al.*, *Nat. Phys.* **9**, 235–241 (2013).
43. J. C. Halimeh, A. Wöllert, I. McCulloch, U. Schollwöck, T. Barthel, *Phys. Rev. A* **89**, 063603 (2014).
44. G. Misguich, K. Mallick, P. L. Krapivsky, *Phys. Rev. B* **96**, 195151 (2017).
45. O. Gamayun, Y. Miao, E. Ilievski, *Phys. Rev. B* **99**, 140301 (2019).
46. B. Ye, F. Machado, C. D. White, R. S. K. Mong, N. Y. Yao, *Phys. Rev. Lett.* **125**, 030601 (2020).
47. C. D. White, M. Zaletel, R. S. K. Mong, G. Refael, *Phys. Rev. B* **97**, 035127 (2018).
48. J. De Nardis, S. Gopalakrishnan, E. Ilievski, R. Vasseur, *Phys. Rev. Lett.* **125**, 070601 (2020).
49. Y. Tang *et al.*, *Phys. Rev. X* **8**, 021030 (2018).
50. M. A. Nichols *et al.*, *Science* **363**, 383–387 (2019).
51. J. De Nardis, S. Gopalakrishnan, R. Vasseur, B. Ware, *Phys. Rev. Lett.* **127**, 057201 (2021).
52. F. Weiner, P. Schmitteckert, S. Bera, F. Evers, *Phys. Rev. B* **101**, 045115 (2020).
53. M. Fava, B. Ware, S. Gopalakrishnan, R. Vasseur, S. A. Parameswaran, *Phys. Rev. B* **102**, 115121 (2020).
54. M. Prähofer, H. Spohn, *Phys. Rev. Lett.* **84**, 4882–4885 (2000).
55. F. Family, T. Vicsek, *J. Phys. Math. Gen.* **18**, L75–L81 (1985).
56. D. Wei *et al.*, Quantum gas microscopy of Kardar-Parisi-Zhang superdiffusion, version 1.0, Edmond (2022); https://doi.org/10.17617/3.8w.

ACKNOWLEDGMENTS

We gratefully acknowledge discussions with J. Moore, R. Vasseur, and M. Zaletel. We thank K. Takeuchi, T. Prosen, and H. Spohn for comments on the manuscript. **Funding:** We acknowledge funding from the Max Planck Society (MPG), the European Union (PASQuanS grant no. 817482), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2111 – 390814868. B.Y., F.M., J.K., and N.Y.Y. acknowledge support from the ARO (grant no. W911NF-21-1-0262) and through the MURI program (W911NF-20-1-0136). J.R. acknowledges funding from the Max Planck Harvard Research Center for Quantum Optics. S.G. acknowledges support from the NSF (DMR-1653271). N.Y.Y. acknowledges support from the David and Lucile Packard Foundation and the W. M. Keck Foundation. **Author contributions:** All authors contributed substantially to the work presented in this manuscript. D.W., A.R.-A., K.S., and S.H. acquired the data and maintained the experimental apparatus. D.W. analyzed the data. B.Y., F.M., J.K., S.G., and N.Y.Y. performed the numerical and analytical calculations. I.B. and J.Z. supervised the study. All authors worked on the interpretation of the data and contributed to writing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data shown in the main text and supplementary materials and the code associated with the numerical simulations are available from the Edmond repository of the Max Planck Society (*56*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk2397
Supplementary Text
Figs. S1 to S18
Tables S1 and S2
References (*57*–*63*)

Submitted 30 June 2021; accepted 10 March 2022
10.1126/science.abk2397

QUANTUM SIMULATION

Observing emergent hydrodynamics in a long-range quantum magnet

M. K. Joshi[1], F. Kranzl[1,2], A. Schuckert[3,4], I. Lovas[3,4], C. Maier[1,5], R. Blatt[1,2], M. Knap[3,4]*, C. F. Roos[1,2]*

Identifying universal properties of nonequilibrium quantum states is a major challenge in modern physics. A fascinating prediction is that classical hydrodynamics emerges universally in the evolution of any interacting quantum system. We experimentally probed the quantum dynamics of 51 individually controlled ions, realizing a long-range interacting spin chain. By measuring space-time–resolved correlation functions in an infinite temperature state, we observed a whole family of hydrodynamic universality classes, ranging from normal diffusion to anomalous superdiffusion, that are described by Lévy flights. We extracted the transport coefficients of the hydrodynamic theory, reflecting the microscopic properties of the system. Our observations demonstrate the potential for engineered quantum systems to provide key insights into universal properties of nonequilibrium states of quantum matter.

In quantum many-body systems at equilibrium, the concept of universality asserts that microscopic details do not influence the nature of the emergent quantum phases of matter and their transitions. Rather, symmetries and topology determine the essential macroscopic properties. By contrast, all scales, from low to high energies, are a priori relevant for quantum systems that are driven far from their thermal equilibrium. Recent experimental progress in engineering coherent and interacting quantum systems made it possible to create and explore exotic nonequilibrium states, which can exhibit unconventional relaxation dynamics (*1*–*6*), dynamical phases (*7*–*12*), and transitions between them (*13*, *14*).

Despite this wealth of observed quantum phenomena, a common anticipation is that classical hydrodynamics of a few conserved quantities emerges universally for any complex quantum system, as strong interactions entangle and effectively mix local degrees of freedom (*15*, *16*). However, verifying this assumption, and furthermore determining the nonuniversal transport coefficients of the emergent hydrodynamic theory for specific systems, is challenging. Recently, enormous efforts have been devoted to detect hydrodynamic transport in quantum gases (*17*–*20*) and condensed matter systems (*21*–*24*). Although transport is generally expected to be diffusive, a variety of largely unexplored classes of hydrodynamics have been predicted theoretically, including anomalous subdiffusive (*25*, *26*) and superdiffusive transport (*27*–*29*).

In this work, we experimentally probed the dynamics of a long-range quantum magnet, realized in a quantum simulator of 51 individually controlled ions. We developed a protocol to measure space- and time-resolved spin correlations in an engineered infinite temperature state, enabling us to experimentally establish that hydrodynamics emerges in the nonequilibrium quantum state. By tuning the long-range character of the interactions, we observed a whole family of hydrodynamical universality classes, ranging from normal diffusion to anomalous superdiffusion.

[1]Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Technikerstraße 21a, 6020 Innsbruck, Austria. [2]Institut für Experimentalphysik, Universität Innsbruck, Technikerstraße 25, 6020 Innsbruck, Austria. [3]Department of Physics and Institute for Advanced Study, Technical University of Munich, 85748 Garching, Germany. [4]Munich Center for Quantum Science and Technology (MCQST), Schellingstraße 4, 80799 München, Germany. [5]AQT, Technikerstraße 17, 6020 Innsbruck, Austria.
***Corresponding author. Email: michael.knap@ph.tum.de (M.K.); christian.roos@uibk.ac.at (C.F.R.)**

Engineering a long-range quantum magnet

The (pseudo-)spins of our quantum system are realized with two electronic states of the $^{40}Ca^{+}$ ion: $\left|S_{\frac{1}{2}}, m=+\frac{1}{2}\right\rangle$ as $|\downarrow\rangle$ and $\left|D_{5/2}, m=+5/2\right\rangle$

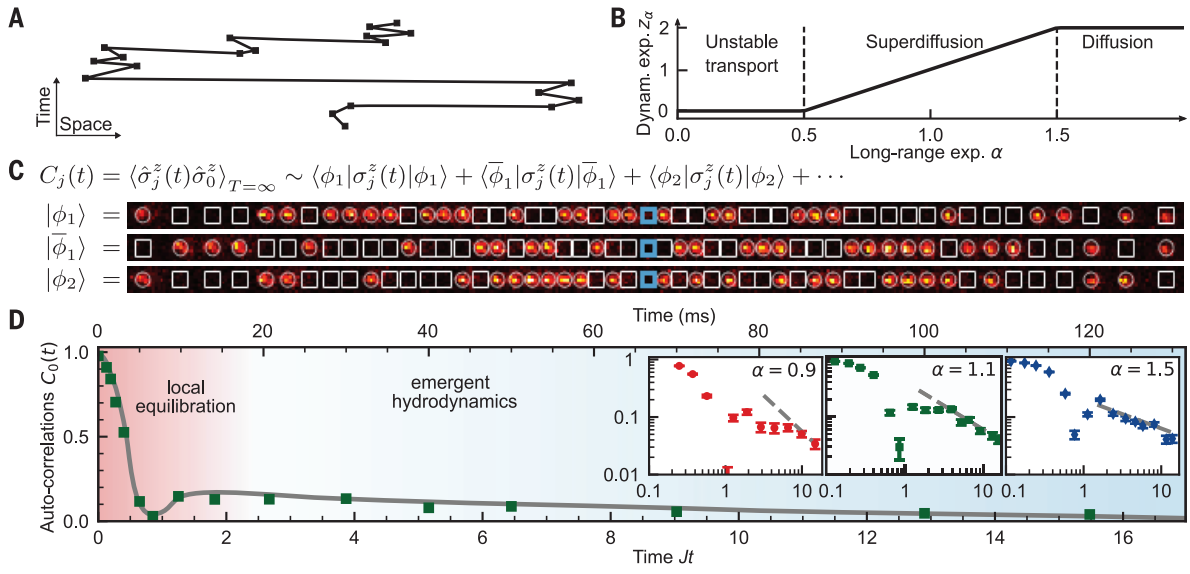


Fig. 1. Emergent hydrodynamics in a long-range quantum magnet. (**A**) In our system, the dynamics of a locally created spin excitation is effectively governed by a hydrodynamic theory of Lévy flights, which are random walks with step sizes drawn from a distribution with algebraic tails. This occasionally leads to long-distance jumps of the excitation. (**B**) The family of hydrodynamic universality classes characterized by a dynamical exponent z_α and a scaling function can be tuned by the power-law exponent α of the spin-spin interactions. (**C**) We measured infinite-temperature correlations by averaging over initial product states and preparing the central ion deterministically in the same state (blue box) [(*32*), section 2]. Picture of three exemplary initial states in a chain of $^{40}Ca^+$ ions ($|\uparrow\rangle$, dark spots; $|\downarrow\rangle$, bright spots). White squares and circles indicate the intended preparation of $|\uparrow\rangle$ and $|\downarrow\rangle$, respectively, achieved with 99% fidelity per ion. (**D**) Measured autocorrelations (green squares) for 51 ions and α = 1.1 (gray line is a guide to the eye). Error bars, denoting the standard error of the mean, are smaller than the symbols [(*32*), section 5]. At short times (red shading), the spin excitation quickly relaxes to a local equilibrium state. At late times (blue shading), global conservation laws constrain the relaxation of spin excitations, leading to a slow power-law decay of the autocorrelations. (Insets) Autocorrelations on a double logarithmic scale for different values of α highlighting the tunable transport. Gray dashed lines are power laws with the predicted exponent from Lévy flights. Here, J = 248, 129, and 116 rad/s for α = 0.9 (51 ions), α = 1.1 (51 ions), and α = 1.5 (25 ions), respectively.

as $|\uparrow\rangle$. The quantum state of individual ions is controlled by a tightly focused, steerable laser beam capable of addressing any ion in the string, in conjunction with a laser beam that collectively interacts with the ions. A two-tone laser field realizes approximately power-law decaying Ising interactions between the (pseudo-)spins by off-resonantly coupling motional and electronic degrees of freedom of the ion chain. Application of a strong transverse field energetically penalizes spin nonconserving contributions (*30*). The effective dynamics are then described by the long-range XY model

$$\hat{H} = \sum_{i<j} \frac{J}{|i-j|^\alpha}\left(\hat{\sigma}_i^+\hat{\sigma}_j^- + \hat{\sigma}_i^-\hat{\sigma}_j^+\right) \quad (1)$$

where J and α determine the strength and the range of the spin-spin interaction, respectively (*31*). They can be tuned by varying the amplitude and frequency of the two-tone laser field. $\hat{\sigma}_i^+$ and $\hat{\sigma}_i^-$ are the raising and lowering spin operators acting on the ith spin, respectively. Further experimental details are given in the supplementary materials [(*32*), sections 1 to 4].

Theoretical expectations for late-time dynamics

Conservation laws determine the macroscopic late-time dynamics of quantum systems. In the long-range XY model, the global U(1) symmetry implies the conservation of the total magnetization $\sum_j \hat{\sigma}_j^z$. This constrains the relaxation of spin excitations. The long-range couplings enable transport of spin excitations over many lattice sites. As a consequence, at late times the effective dynamics are expected to be governed by classical Lévy flights, which describe random walks with long-distance jumps (Fig. 1A) (*29*, *33*). The probability for a jump of distance ℓ in the Lévy flight is set by the power-law couplings of the long-range XY model as $\sim J^2/\ell^{2\alpha}$ [(*32*), section 8]. These long-distance jumps become irrelevant when the variance of the jump probability is finite. In our model, this is the case for α > 3/2, implying conventional diffusion [(*32*), section 8]. However, the variance diverges for α ≤ 3/2, giving rise to a breakdown of conventional diffusion. The hydrodynamics of our system is therefore expected to be described by the anomalous diffusion equation

$$\partial_t \left\langle \hat{\sigma}_j^z \right\rangle = D_\alpha \nabla^{z_\alpha} \left\langle \hat{\sigma}_j^z \right\rangle \quad (2)$$

where D_α denotes the associated transport coefficient and z_α is the dynamical exponent [(*32*), section 8]. The dynamical exponent z_α relates space and time and hence characterizes transport. For α > 3/2, transport is diffusive, and space and time are related by $z_\alpha = 2$. For 1/2 ≤ α ≤ 3/2, we predict $z_\alpha = 2\alpha - 1$ (*29*), leading to anomalous superdiffusion. For α < 1/2, the spins are so strongly connected that the lattice geometry becomes irrelevant, and a regime with $z_\alpha = 0$ is entered. Therefore, transport is strongly modified by the long-range character of the interactions. The predicted dynamical phase diagram is shown in Fig. 1B.

Experimental signatures of emergent hydrodynamics

Hydrodynamic transport is most directly probed by creating a spin excitation at time $t = 0$ in the center of the chain and tracking how it propagates in space and time, as measured by the unequal-time correlation function

$$C_j(t) = \left\langle \hat{\sigma}_j^z(t)\hat{\sigma}_0^z \right\rangle_{T=\infty} \quad (3)$$

Over the course of the quantum dynamics, the initial excitation has to scatter off a highly excited background to quickly reach the hydrodynamic regime. Therefore, we measured $C_j(t)$ at infinite temperatures T with magnetization ~0.

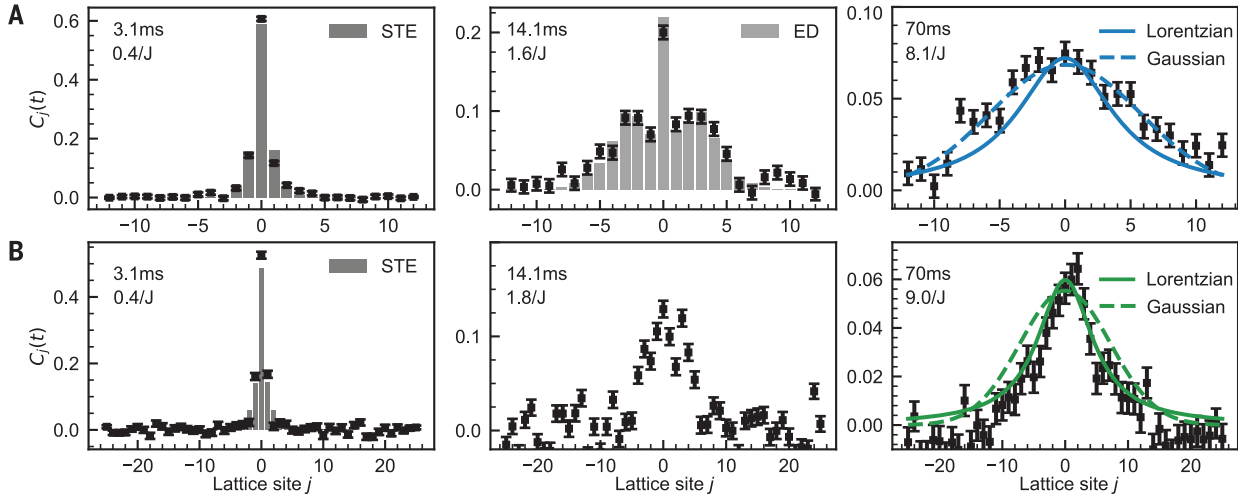


Fig. 2. Spatial correlation profiles. (**A** and **B**) Spatially resolved correlations for (A) $\alpha = 1.5$ for 25 ions and (B) $\alpha = 1.1$ for 51 ions. The spatial correlations are measured at (left) short times, (middle) intermediate times, and (right) in the hydrodynamic late-time regime. STE, analytic short-time expansion; ED, exact diagonalization accessible only for the shorter chain of 25 ions. In the hydrodynamic regime (right), the measured profiles are compared with predictions from Lévy flights. The spatial profile in (A) is compatible with a Gaussian (dashed) and in (B) with a Lorentzian (solid), as supported by the reduced χ^2 values of the fit: (A) $\chi^2_L = 3.9$, $\chi^2_G = 1.3$; (B) $\chi^2_L = 1.4$, $\chi^2_G = 4.8$ (obtained by fitting over the central 27 sites). Here, χ^2_L and χ^2_G are χ^2 values for the Lorentzian and Gaussian fits, respectively.

Measuring spin correlations at different times and preparing an infinite temperature state are both difficult in isolated, engineered quantum systems. We overcame these challenges by expressing the infinite temperature expectation value as an equally weighted trace over product states $|\phi\rangle$ in the $\hat{\sigma}^z$ basis [(*32*), section 10]. When preparing the central spin in the same polarization, we have $\hat{\sigma}^z_0|\phi\rangle = +|\phi\rangle$ (Fig. 1C). The correlation function can then be directly evaluated as $C_j(t) \sim \langle\phi_1|\hat{\sigma}_j z(t)|\phi_1\rangle + \langle\phi_2|\hat{\sigma}_j z(t)|\phi_2\rangle + \ldots$. It can be accurately obtained by sampling a finite number of initial product states [(*32*), section 10]. Yet for a comparatively small and experimentally accessible number of initial states, large statistical fluctuations are expected in the measured correlations. We removed these fluctuations by sampling pairs of conjugate product states, $|\phi\rangle$ and $|\bar{\phi}\rangle$, where in the second configuration all spins are flipped except for the central one. For each pair of product states, initial correlations are unity in the center of the system and zero elsewhere, reproducing directly this property of the full trace (Fig. 1C, first two initial states). With this procedure, convergence is achieved even for a small number of initial product states [(*32*), section 10]. For $\alpha = 0.9$ and 1.1 and for $\alpha = 1.5$, respectively, we created 60 and 120 initial product state configurations, each of which was realized, evolved, and measured 50 to 200 times.

The autocorrelation function $C_0(t)$ determines the residual excitation in the center of the chain. We measured $C_0(t)$ for 51 ions (Fig. 1D). The relaxation dynamics occurred in multiple stages. A local equilibrium was reached already after a few collisions, with the abundant excitations of the infinite temperature state. As a consequence, at short times the autocorrelation exhibited a rapidly damped oscillation. At later times, the system entered the hydrodynamic regime and eventually approached a global equilibrium through the slow rearrangement of spin excitations constrained by the conserved magnetization. The longest time to which we could probe transport was limited by the saturation of the correlation function set by the inverse system size. The power-law decay of the correlations became manifest on a double logarithmic scale (Fig. 1D, insets). Our experimental data are consistent with the hydrodynamic theory of Eq. 2, which predicts a power-law exponent of $-1/z_\alpha$ (Fig. 1D, insets, dashed gray lines). Hence, exploiting the experimental control over the long-range exponent α enabled us to tune the transport from normal diffusion to anomalous superdiffusion.

We obtained a more stringent test of the emergent hydrodynamics from the full spatial correlation profile (Fig. 2). We realized the long-range spin model for $\alpha = 1.5$ with 25 ions (Fig. 2A) and for $\alpha = 1.1$ with 51 ions (Fig. 2B). At early times (3.1 ms), the quantum dynamics are well-described by an analytic short-time expansion of the equations of motion [(*32*), section 9]. At intermediate times (14.1 ms), the excitation starts spreading through the system, but some quantum coherence remains, indicated by the spatial oscillations. For 25 ions, the measured dynamics compares well with results obtained from exact diagonalization, demonstrating the coherence of our quantum system [(*32*), section 7]. For the 51 ion chain, a comparison to exact diagonalization was not possible because of the exponential growth of the Hilbert space, which reached a dimension of $2^{51} \approx 10^{15}$.

The quantum dynamics effectively coarsens the system during the time evolution as spatial oscillations with longer wavelengths start to dominate. At the latest times shown (70 ms), interactions have averaged out quantum interference patterns, and the hydrodynamic regime is entered. This became even more apparent in the space-time correlations (Fig. 3A). By contrast, a single excitation on top of a spin-polarized state cannot scatter, and the associated correlations exhibit coherent space-time oscillation patterns instead (Fig. 3B) (*30*).

The theory of Lévy flights predicts the following scaling form for the spatiotemporal profiles

$$C_j(t) = (D_\alpha t)^{-\frac{1}{z_\alpha}} F_\alpha\left[\frac{|j|}{(D_\alpha t)^{1/z_\alpha}}\right] \qquad (4)$$

where F_α is given by the family of stable symmetric distributions (*33*). The scaling between space and time entering the distribution F_α can be deduced directly from the generalized hydrodynamic equation Eq. 2. The distribution F_α cannot be expressed in terms of elementary functions, except for $\alpha = 3/2$, where a Gaussian indicates conventional diffusion, and $\alpha = 1$, where a Lorentzian appears.

The measured hydrodynamic profiles are compatible with a Lorentzian for $\alpha = 1.1$ ($F_{\alpha=1.1}$ is still very close to a Lorentzian)

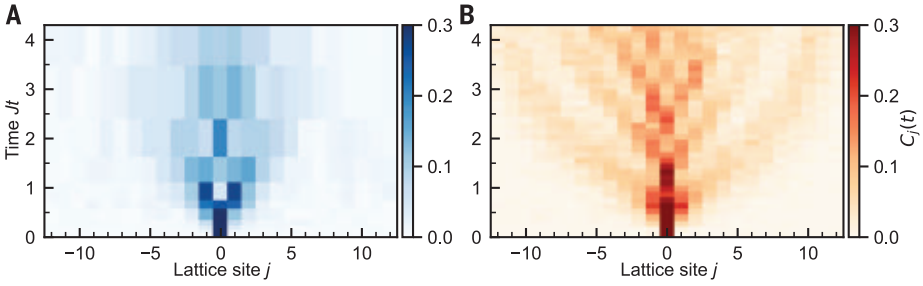


Fig. 3. Contrasting the infinite temperature background with the spin polarized background. (**A**) The deterministically prepared excitation of the central ion strongly interacts with all the other excitations of the infinite temperature background and slowly spreads through the system following the laws of classical hydrodynamics. (**B**) By contrast, a single excitation on top of the fully down-polarized state $|\downarrow \ldots \downarrow\uparrow\downarrow \ldots\downarrow\rangle$ has no other excitations to scatter off and therefore spreads freely, exhibiting quantum interference patterns. Data are measured for 25 ions with power-law exponent $\alpha = 1.5$.

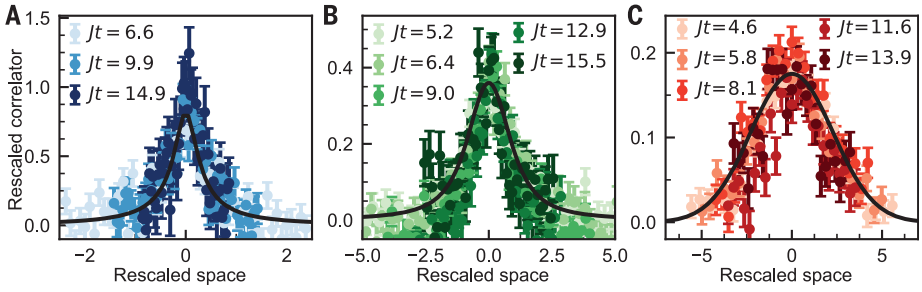


Fig. 4. Extracting transport coefficients. At late times, the correlations exhibit a self-similar scaling, relating space and time with a microscopic transport coefficient D_α. Shown are rescaled correlations $\tau_x C_j(t)$ as a function of rescaled space j/τ_x, where $\tau_x = (Jt)^{1/z_\alpha}$. (**A**) $\alpha = 0.9$, 51 ions. (**B**) $\alpha = 1.1$, 51 ions. (**C**) $\alpha = 1.5$, 25 ions. Correlations are shown for times $Jt > 5$; darker colors correspond to later times.

and with a Gaussian for $\alpha = 1.5$ (χ^2 analysis is provided in Fig. 2, right), which is in agreement with Eq. 4.

Determining scaling exponents and coefficients

Close to equilibrium phase transitions, correlation functions show self-similarity as a function of the distance to the critical point and can be characterized by a set of universal scaling exponents as well as scaling functions (*34*). Hydrodynamics predicts a self-similar scaling of correlations as a function of time, characterized by the dynamical scaling exponent z_α and the scaling function F_α (*35*). Whereas z_α and F_α solely depend on the asymptotic decay of the interactions and can be predicted from purely hydrodynamic reasoning, the transport coefficient D_α is not universal and depends on the full quantum many-body spectrum. It is therefore challenging to make predictions of D_α by using analytical or numerical methods. To verify Eq. 4, we rescaled $C_j(t)$ and space j as shown in Fig. 4, leading to a scaling collapse. By fitting the stable symmetric distributions to the collapsed data, we obtained a transport coefficient of $D_\alpha/J = 0.5^{+0.2}_{-0.1}$, $0.8^{+0.3}_{-0.2}$, and $2.6^{+0.9}_{-0.7}$ for $\alpha = 0.9$, 1.1, and 1.5, respectively [(*32*), section 6]. From exact diagonalization, we obtained $D_\alpha/J = 1.9^{+0.7}_{-0.5}$ for $\alpha = 1.5$.

Discussion

Large, coherent quantum systems with local control provide key insights into fundamental properties of nonequilibrium quantum states. In our experiment, we measured the emergent macroscopic hydrodynamics in an infinite temperature state of a strongly interacting spin chain—a regime that is notoriously challenging to describe by controlled analytical or numerical calculations. By measuring the full spatiotemporal profile of the hydrodynamic scaling functions, we experimentally established a tunable family of transport that ranges from conventional diffusion to anomalous superdiffusion.

One prospect is to investigate the generality of the dynamical phase diagram (Fig. 1B) for other models with long-range interactions that conserve the total magnetization or charge. An example could be the long-range Heisenberg model, which can be realized in our system with suitably designed Floquet protocols (*36*). Moreover, our tools for measuring high-temperature correlations can be readily applied to other quantum devices with local control, including quantum gas microscopes, Rydberg atom arrays, and superconducting qubits. Diffusion constants measured in controlled synthetic quantum matter could serve as a benchmark for newly developed numerical methods and for comparing different experimental platforms that realize the same model. Creating finite temperature states in the vicinity of a thermal phase transition with synthetic quantum matter would enable the study of emergent dynamic critical phenomena, describing a modified hydrodynamics in which additional conservation laws are introduced by the order parameter (*37*). Emergent hydrodynamics could also be tested for different initial states, such as certain product states (*15*).

Note added in proof: During the completion of this manuscript, we became aware of related work demonstrating superdiffusive transport in an integrable Heisenberg chain with nearest-neighbor superexchange interactions (*38*).

REFERENCES AND NOTES

1. M. Gring *et al.*, *Science* **337**, 1318–1322 (2012).
2. J. G. Bohnet *et al.*, *Science* **352**, 1297–1301 (2016).
3. Y. Tang *et al.*, *Phys. Rev. X* **8**, 021030 (2018).
4. H. Bernien *et al.*, *Nature* **551**, 579–584 (2017).
5. E. Guardado-Sanchez *et al.*, *Phys. Rev. X* **10**, 011042 (2020).
6. S. Scherg *et al.*, *Nat. Commun.* **12**, 4490 (2021).
7. M. Schreiber *et al.*, *Science* **349**, 842–845 (2015).
8. J. Smith *et al.*, *Nat. Phys.* **12**, 907–911 (2016).
9. A. M. Kaufman *et al.*, *Science* **353**, 794–800 (2016).
10. T. Brydges *et al.*, *Science* **364**, 260–263 (2019).
11. M. Prüfer *et al.*, *Nature* **563**, 217–220 (2018).
12. S. Erne, R. Bücker, T. Gasenzer, J. Berges, J. Schmiedmayer, *Nature* **563**, 225–229 (2018).
13. P. Jurcevic *et al.*, *Phys. Rev. Lett.* **119**, 080501 (2017).
14. J. Zhang *et al.*, *Nature* **551**, 601–604 (2017).
15. J. Lux, J. Müller, A. Mitra, A. Rosch, *Phys. Rev. A* **89**, 053608 (2014).
16. A. Bohrdt, C. B. Mendl, M. Endres, M. Knap, *New J. Phys.* **19**, 063001 (2017).
17. C. Cao *et al.*, *Science* **331**, 58–61 (2011).
18. A. Sommer, M. Ku, G. Roati, M. W. Zwierlein, *Nature* **472**, 201–204 (2011).
19. U. Schneider *et al.*, *Nat. Phys.* **8**, 213–218 (2012).
20. P. T. Brown *et al.*, *Science* **363**, 379–382 (2019).
21. D. A. Bandurin *et al.*, *Science* **351**, 1055–1058 (2016).
22. J. Crossno *et al.*, *Science* **351**, 1058–1061 (2016).
23. P. J. W. Moll, P. Kushwaha, N. Nandi, B. Schmidt, A. P. Mackenzie, *Science* **351**, 1061–1064 (2016).
24. C. Zu *et al.*, *Nature* **597**, 45–50 (2021).
25. A. Gromov, A. Lucas, R. M. Nandkishore, *Phys. Rev. Res.* **2**, 033124 (2020).
26. J. Feldmeier, P. Sala, G. De Tomasi, F. Pollmann, M. Knap, *Phys. Rev. Lett.* **125**, 245303 (2020).
27. M. Ljubotina, M. Žnidarič, T. Prosen, *Nat. Commun.* **8**, 16117 (2017).
28. V. B. Bulchandani, S. Gopalakrishnan, E. Ilievski, *J. Stat. Mech.* **2021**, 084001 (2021).
29. A. Schuckert, I. Lovas, M. Knap, *Phys. Rev. B* **101**, 020416 (2020).
30. P. Jurcevic *et al.*, *Nature* **511**, 202–205 (2014).
31. C. Monroe *et al.*, *Rev. Mod. Phys.* **93**, 025001 (2021).
32. Materials and methods are available as supplementary materials.
33. V. Zaburdaev, S. Denisov, J. Klafter, *Rev. Mod. Phys.* **87**, 483–530 (2015).
34. J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge Univ. Press, 1996).
35. D. Forster, *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions* (CRC Press, 1975).
36. S. Birnkammer, A. Bohrdt, F. Grusdt, M. Knap, arXiv:2012.09185 [cond-mat.quant-gas] (2020).

37. P. C. Hohenberg, B. I. Halperin, *Rev. Mod. Phys.* **49**, 435–479 (1977).
38. D. Wei *et al.*, *Science* **376**, 716–720 (2022).
39. M. K. Joshi *et al.*, Observing emergent hydrodynamics in a long-range quantum magnet, Zenodo (2022); https://doi.org/10.5281/zenodo.6256977.

ACKNOWLEDGMENTS

Funding: We acknowledge support from the Technical University of Munich–Institute for Advanced Study, funded by the German Excellence Initiative and the European Union FP7 under grant agreement 291763; the Max Planck Gesellschaft (MPG) through the International Max Planck Research School for Quantum Science and Technology (IMPRS-QST); the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under Germany's Excellence Strategy–EXC–2111–390814868; as well as the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 817482, from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 851161, no. 741541, and no. 771537), the Institut für Quanteninformation GmbH, and the Austrian Science Fund through the SFB BeyondC (F7110). **Author contributions:** M.K.J., C.M., M.K., and C.F.R. devised the research. A.S., I.L., and M.K. developed the theoretical protocols. M.K.J., F.K., C.M., R.B., and C.F.R. contributed to the experimental setup. M.K.J. and F.K. performed the experiments. M.K.J., A.S., I.L., M.K., and C.F.R. analyzed the data, and M.K.J., A.S., and C.F.R. carried out numerical simulations. M.K.J., A.S., M.K., and C.F.R. wrote the manuscript. All authors contributed to the discussion of the results and the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data analysis and simulation codes along with the experimental data are available on Zenodo (*39*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk2400
Materials and Methods
Supplementary Text
Figs. S1 to S9
References (*40*–*46*)

Submitted 30 June 2021; accepted 10 March 2022
10.1126/science.abk2400

NEUROSCIENCE

Paradoxical somatodendritic decoupling supports cortical plasticity during REM sleep

Mattia Aime[1,2], Niccolò Calcini[1,2], Micaela Borsa[1,2], Tiago Campelo[1,2], Thomas Rusterholz[1,2], Andrea Sattin[3], Tommaso Fellin[3], Antoine Adamantidis[1,2]*

Rapid eye movement (REM) sleep is associated with the consolidation of emotional memories. Yet, the underlying neocortical circuits and synaptic mechanisms remain unclear. We found that REM sleep is associated with a somatodendritic decoupling in pyramidal neurons of the prefrontal cortex. This decoupling reflects a shift of inhibitory balance between parvalbumin neuron–mediated somatic inhibition and vasoactive intestinal peptide–mediated dendritic disinhibition, mostly driven by neurons from the central medial thalamus. REM-specific optogenetic suppression of dendritic activity led to a loss of danger-versus-safety discrimination during associative learning and a lack of synaptic plasticity, whereas optogenetic release of somatic inhibition resulted in enhanced discrimination and synaptic potentiation. Somatodendritic decoupling during REM sleep promotes opposite synaptic plasticity mechanisms that optimize emotional responses to future behavioral stressors.

Rapid eye movement [(REM) or paradoxical] sleep is characterized by a specific activation of neuronal populations that encode emotion and reward in animals (*1*–*3*) and humans (*4*–*6*). This REM sleep–specific activity is associated with molecular and cellular fingerprints of enhanced synaptic plasticity, including up-regulated gene expression, long-term potentiation (LTP) (*7*–*9*), and heightened dendritic activity (*10*, *11*) in the cortex and hippocampus that ultimately consolidates memory traces (*12*, *13*). This is in contrast to the quiescence of brain structures, including the prefrontal cortex (PFC), that is associated with the pruning of dendritic spines (*11*, *14*), synaptic weakening (*15*), and forgetting during REM sleep (*16*–*18*). These apparent controversies are the result of the heterogeneity of (sub)cellular activity in limbic (*19*, *20*) or cortical structures (*4*, *11*, *21*–*23*), limited cellular resolution of human brain imaging technologies (*4*–*6*), and the lack of selectivity of REM sleep–deprivation procedures (*8*, *24*). In this work, we investigated the mechanisms encoding emotional memories during REM sleep within dorsal PFC (dPFC) microcircuits—a critical brain structure for executive functions (*25*) and discrimination of danger from safety (*26*).

[1]Zentrum für Experimentelle Neurologie, Department of Neurology, Inselspital University Hospital Bern, Bern, Switzerland. [2]Department of Biomedical Research, University of Bern, Bern, Switzerland. [3]Optical Approaches to Brain Function Laboratory, Istituto Italiano di Tecnologia, Genova, Italy.
***Corresponding author. Email: antoine.adamantidis@dbmr.unibe.ch**

Decoupling of somatodendritic activities in cortical neurons during REM sleep

We first dissected dPFC microcircuit dynamics during REM sleep in head-restrained sleeping mice using simultaneous two-photon calcium imaging and electrophysiological recordings from chronic electroencephalogram (EEG) and electromyogram (EMG) electrodes (Fig. 1A). An adeno-associated virus encoding the calcium indicator GCaMP6 under the calmodulin kinase II promoter (AAV1-CaMKII-GCaMP6) was stereotactically injected into the dPFC area. The resulting GCaMP6 expression was specific to pyramidal (PYR) neurons of superficial layers (L2/3) (Fig. 1, A and B). Simultaneous recordings revealed a significant decrease of PYR somatic activity during REM sleep as compared with non-REM (NREM) sleep or wakefulness (Fig. 1, C and D; movie S1; and fig. S1; see materials and methods for data analysis and motion artifact detection). This decrease was consistent across all animals tested (fig. S2) and was confirmed using deconvolution analytical methods (materials and methods and fig. S3, A to D). Although few PYR somas remained active during REM sleep (8%), a substantial set of PYR cells was preferentially active during either NREM sleep (31%) or wakefulness (61%) (Fig. 1E). Distribution of single–PYR cell $\Delta F/F_0$ values (where $\Delta F/F_0$ is the ratio of the change in fluorescence to the baseline fluorescence) indicated low activity during REM sleep as compared with NREM sleep and wakefulness (Fig. 1F). REM sleep was associated with a significant decrease in the frequency of calcium (Ca^{2+}) events compared with those in other states (~18%; versus ~33% in NREM sleep or ~49% in wakefulness) (Fig. 1, G and H), although the integrals and amplitudes of Ca^{2+} events were similar across states (fig. S4).

We next tested whether apical dendrites of L2/3 dPFC PYR neurons were similarly modulated during REM sleep in mice cotransduced with an AAV1-Syn-flex-GCaMP6 and an AAV1-CaMKII-Cre to obtain a sparse labeling of dendrites (materials and methods; Fig. 1, I and J; and movie S1). Spatial propagation of GCaMP6s activity along individual dendrites was quantified for all GCaMP6s-expressing dendrites. Each identified branch was segmented into 0.5-μm-wide rectangular regions of interest (ROIs), and $\Delta F/F_0$ traces of each ROI were extracted (materials and methods and Fig. 1, K and L). In contrast to PYR somas, PYR dendrites were highly active during REM sleep (Fig. 1M), as shown by the significant increase of the frequency of events spreading along the dendrites (>20 μm) compared with events confined to small portions of the dendritic branches (<20 μm). The average propagation of events along the dendrites was also higher during REM sleep (Fig. 1, N and O), providing evidence for a REM sleep–specific disinhibition of apical dendrites concomitant with a tonic inhibition of PYR somas.

Shift in excitation/inhibition balance during REM sleep in cortical microcircuits

To investigate the excitation/inhibition (E/I) balance during REM sleep, we recorded the

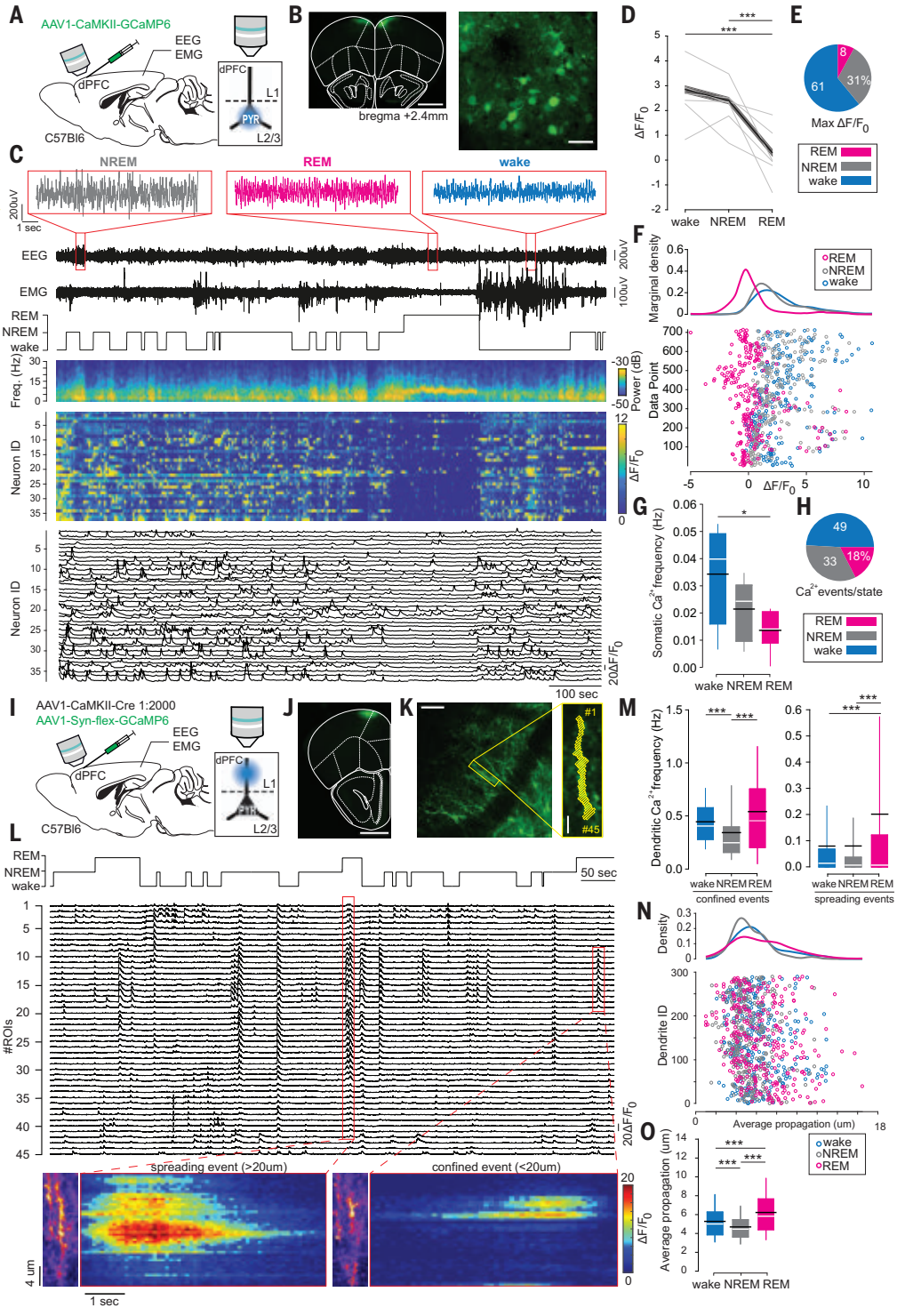


Fig. 1. Somatodendritic decoupling during REM sleep in dPFC PYR neurons. (**A**) Schematic of the experimental procedure. (**B**) (Left) Expression profile of L2/3 GCaMP6 PYR neurons. Scale bar, 1 mm. (Right) GCaMP6 PYR neurons. Scale bar, 20 μm. (**C**) EEG [red squares are magnifications of NREM sleep (gray), REM sleep (fuchsia), and wake (light blue) traces], EMG, hypnogram, color-coded power spectrum density, and color-coded and raw $\Delta F/F_0$ activity profiles of a population of PYR somas. Freq., frequency. (**D**) Mean ± SEM (black line) $\Delta F/F_0$ of PYR somas (n = 238 neurons; m = 6 mice) during wake, NREM, and REM sleep. ***P < 0.001; one-way repeated measures analysis of variance (RM ANOVA) with Tukey post hoc test. (**E**) Percentages of PYR somas with maximal values of $\Delta F/F_0$ in wake (light blue), NREM (gray), and REM (fuchsia) sleep. (**F**) Line (top) and scatter (bottom) plots showing the distribution of $\Delta F/F_0$ values during wake, NREM, and REM sleep. (**G**) Frequency of somatic Ca^{2+} events during wake, NREM, and REM sleep. *P < 0.05; one-way RM ANOVA with Tukey post hoc test. Black lines represent mean values. (**H**) Distribution of somatic Ca^{2+} events in wake, NREM, and REM sleep. (**I**) Schematic of the experimental strategy. (**J**) Expression profile of GCaMP6 in the superficial layers of the dPFC. Scale bar, 1 mm. (**K**) (Left) GCaMP6 PYR dendrites. Scale bar, 20 μm. (Right) Magnification of a segmented dendritic branch. Scale bar, 4 μm. (**L**) (Top) Hypnogram and $\Delta F/F_0$ traces of ROIs from a single dendrite. (Bottom) Color-coded spreading event (left) and confined event (right) from the same dendrite. (**M**) Frequency ± SEM of dendritic Ca^{2+} confined (left) and spreading (right) events (n = 308 dendrites; m = 5 mice) during wake, NREM, and REM sleep. ***P < 0.001; one-way RM ANOVA with Tukey post hoc test. (**N**) Line (top) and scatter (bottom) plots showing the average propagation (in micrometers) of all (confined + spreading) Ca^{2+} events during wake, NREM, and REM sleep. (**O**) Mean propagation of all dendrites during wake, NREM, and REM sleep. ***P < 0.001; one-way RM ANOVA with Tukey post hoc test.

activity of parvalbumin (PV), vasoactive intestinal peptide (VIP), and somatostatin (SST) dPFC interneurons from PV-IRES-Cre, VIP-IRES-Cre, or SST-IRES-Cre mice, respectively, transduced with AAV1-Syn-flex-GCaMP6 in sleeping mice (materials and methods, Fig. 2A, and movie S1). PV interneurons are strategically positioned to exert feed-forward inhibition by innervating principally the PYR somas (*27*). Accordingly, the activity of GCaMP6-expressing PV interneurons was significantly increased during REM sleep (Fig. 2, B and C), an observation that is consistent with the elevated $\Delta F/F_0$ values of GCaMP6-expressing PV neurons during REM sleep (~72% of recorded neurons) (Fig. 2, D, E, and F).

VIP interneurons have been described to preferentially inhibit SST γ-aminobutyric acid (GABA)–releasing cells, which in turn synapse primarily onto the dendritic shaft of PYR neurons (*28*). GCaMP6-expressing VIP

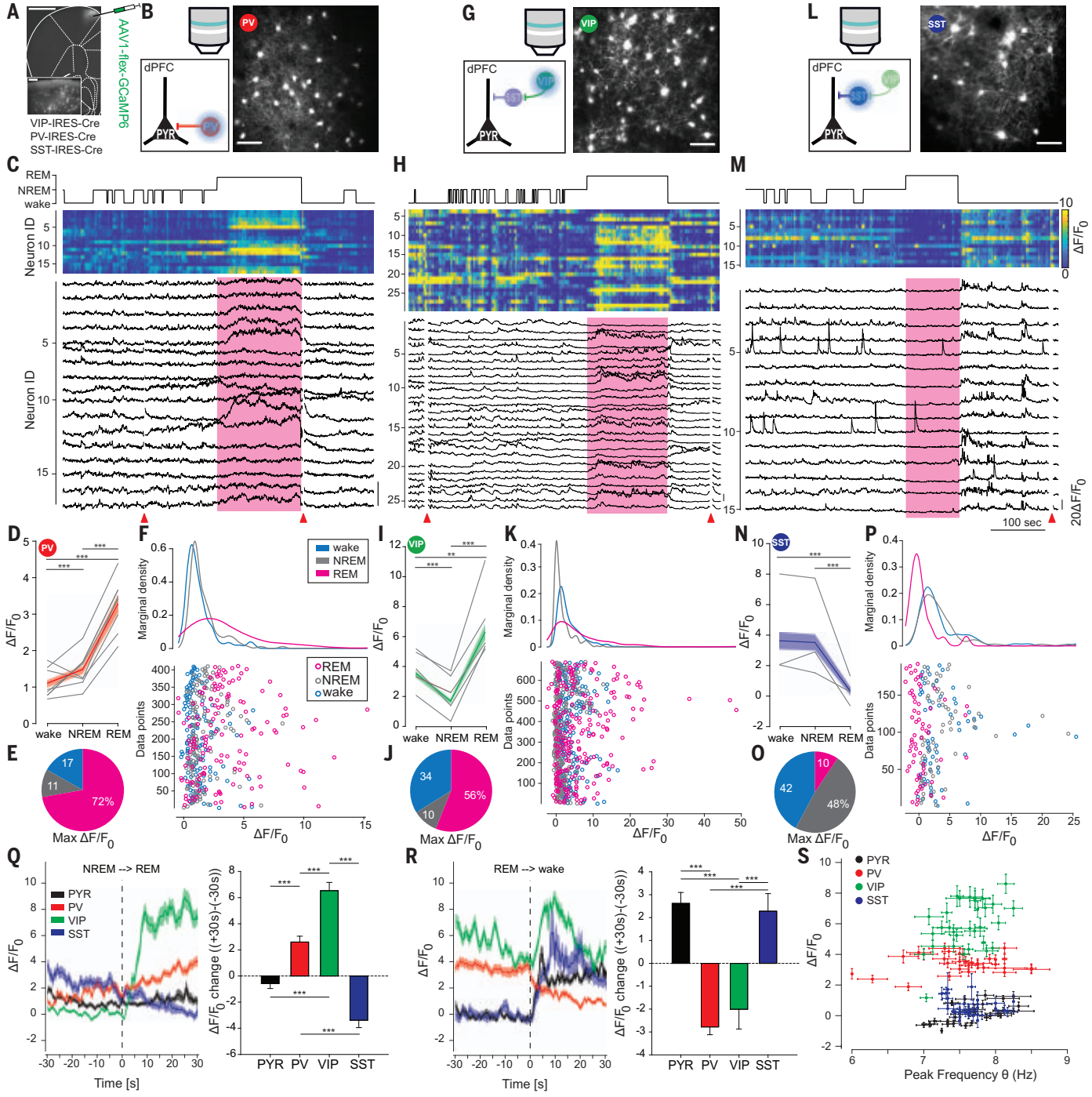


Fig. 2. Shift in inhibitory balance during REM sleep in dPFC circuits. (**A**) Expression profile of GCaMP6 interneurons in the dPFC. Scale bars, 1 mm (top) and 100 μm (bottom). (**B**) (Left) Schematic of the experimental procedure. (Right) GCaMP6 PV neurons. Scale bar, 30 μm. (**C**) Hypnogram and color-coded and raw $\Delta F/F_0$ traces of PV neurons. Red arrowheads indicate motion artifacts. (**D**) Mean ± SEM (red line) $\Delta F/F_0$ of PV neurons (n = 137 neurons; m = 7 mice) during wake, NREM, and REM sleep. ***P < 0.001; one-way RM ANOVA with Tukey post hoc test. (**E**) Percentages of PV neurons with maximal values of $\Delta F/F_0$ in wake (light blue), NREM (gray), and REM (fuchsia) sleep. (**F**) Line (top) and scatter (bottom) plots showing the distribution of PV neurons $\Delta F/F_0$ values during wake, NREM, and REM sleep. (**G**) (Left) Schematic of the experimental procedure. (Right) GCaMP6 VIP neurons. Scale bar, 30 μm. (**H**) Hypnogram and color-coded and raw $\Delta F/F_0$ traces of VIP neurons. Red arrowheads indicate motion artifacts. (**I**) Mean ± SEM (green line) $\Delta F/F_0$ of VIP neurons (n = 230 neurons; m = 5 mice) during wake, NREM, and REM sleep. **P < 0.01; ***P < 0.001; one-way RM ANOVA with Tukey post hoc test. (**J**) Percentages of PV neurons with maximal values of $\Delta F/F_0$ in wake (light blue), NREM (gray), and REM (fuchsia) sleep. (**K**) Line (top) and scatter (bottom) plots showing the distribution of VIP neurons $\Delta F/F_0$ values during wake, NREM, and REM sleep. (**L**) (Left) Schematic of the experimental procedure. (Right) GCaMP6 SST neurons. Scale bar, 30 μm. (**M**) Hypnogram and color-coded and raw $\Delta F/F_0$ traces of SST neurons. Red arrowheads indicate motion artifacts. (**N**) Mean ± SEM (dark blue line) $\Delta F/F_0$ of SST neurons (n = 64 neurons; m = 3 mice) during wake, NREM, and REM sleep. ***P < 0.001; one-way RM ANOVA with Tukey post hoc test. (**O**) Percentages

of SST neurons with maximal values of $\Delta F/F_0$ in wake (light blue), NREM (gray), and REM (fuchsia) sleep. (**P**) Line (top) and scatter (bottom) plots showing the distribution of SST neurons $\Delta F/F_0$ values during wake, NREM, and REM sleep. (**Q**) (Left) Mean ± SEM activity profiles of dPFC neuronal populations during NREM-to-REM sleep transitions (black = PYR, red = PV, green = VIP, dark blue = SST). (Right) $\Delta F/F_0$ change ± SEM across transitions. ***$P < 0.001$; one-way ANOVA with Tukey post hoc test. (**R**) (Left) Mean ± SEM activity profiles of dPFC neuronal populations during the transition from REM sleep to wake. (Right) $\Delta F/F_0$ change ± SEM across transitions. ***$P < 0.001$; one-way ANOVA with Tukey post hoc test. (**S**) Correlation analysis between peak REM sleep theta frequency ± SEM and the respective $\Delta F/F_0$ mean ± SEM of all recorded neurons per session.

interneurons showed a significant enhancement of $\Delta F/F_0$ values during REM sleep, with a large portion of VIP neurons (~56%) displaying maximal activity during REM sleep (Fig. 2, G to K). By contrast, GCaMP6-expressing SST interneurons were mainly silent during REM sleep, except for a small subpopulation (~10%) (Fig. 2, L to P). These results were consistent across animals (fig. S5) and were confirmed with deconvolution analytical methods (fig. S3, E to H, for PV; fig. S3, I to L, for VIP; and fig. S3, M to P, for SST).

The activity of GCaMP6-expressing VIP and PV neurons progressively increased over NREM-to-REM sleep transitions, whereas GCaMP6-expressing PYR somas and SST interneurons both showed opposite responses (Fig. 2Q). Notably, the $\Delta F/F_0$ values of PYR somas during NREM sleep episodes were predictive of the next sleep-wake transition because $\Delta F/F_0$ values were lower in NREM sleep preceding REM sleep than $\Delta F/F_0$ values preceding awakening (fig. S6). Upon awakening, the $\Delta F/F_0$ values of PYR and SST somas significantly increased, whereas those from VIP and PV somas progressively decreased (Fig. 2R). The activity of each population of neurons correlated with tonic (6 to 9 Hz) (Fig. 2S), but not phasic (9 to 13 Hz), REM sleep.

Thalamic control of PV neuron–mediated inhibition of PYR cell somas during REM sleep

We next sought to identify the dPFC synaptic inputs responsible for this REM sleep–specific shift in E/I balance within dPFC microcircuits using a rabies virus–based anatomical mapping and optogenetic circuit probing. To map the synaptic inputs to dPFC PV and VIP neurons, we injected AAV2/1-CMV-Flex-TVA-mCherry-2A-oG and RV-CMV-ENVA-EGFP (EGFP, enhanced green fluorescent protein) (*29*) in PV- and VIP-IRES-Cre mice, respectively (materials and methods). dPFC PV neurons were monosynaptically innervated by neurons located in the central medial thalamus (CM), the agranular retrosplenial cortex (RSA) and granular retrosplenial cortex (RSG), and the primary somatosensory barrel cortex (S1BF) (Fig. 3, A to E). Input cells to VIP neurons were scattered throughout the brain with a few neurons retrogradely labeled in the CM, whereas dPFC PYR neurons were monosynaptically contacted by neurons restricted to the S1 cortical area and the ventrobasalis (VB) nucleus of the thalamus but not the CM. (fig. S7). We then tested for monosynaptic transmission in the CM^{CaMKII}-$dPFC^{PV}$ circuit using in vitro patch-clamp recordings. Genetic targeting of a red fluorescent reporter (mCherry) to PV interneurons was achieved by stereotactic injection of an AA2-EF1-flex-mCherry in the dPFC of PV-IRES-Cre mice, whereas the expression of channelrhodopsin-2 (ChR2) was targeted to excitatory CM neurons by stereotactic infusion of an AAV2-CaMKII-ChR2-eYFP (eYFP, enhanced yellow fluorescent protein) into the CM area (materials and methods and Fig. 3F). Optogenetic activation of ChR2-expressing CM neuron axon terminals in the dPFC elicited reliable excitatory postsynaptic currents (EPSCs) in ~41% of all recorded PV neurons with short latencies (Fig. 3, G to I, and fig. S8A).

We next confirmed that the CM-dPFC pathway was selectively active during REM sleep using two-photon imaging in sleeping mice. To achieve this, AAV1-CaMKII-GCaMP6 virus was stereotactically injected in the CM of C57Bl6 mice and a cranial window was implanted above the dPFC to allow for two-photon imaging of GCaMP6-expressing CM axons in the dPFC (Fig. 3, J to L, and fig. S8B). The activity of CM axons in the PFC showed a significant enhancement during REM sleep (Fig. 3, M and N), with ~77% of all recorded projections displaying maximal $\Delta F/F_0$ values during REM sleep (Fig. 3O). This finding was consistent with the high activity profiles and high state-selectivity of a subset of the CM cell somas (17%) during REM sleep, as shown by two-photon imaging of CM neurons through a microendoscope lens (materials and methods and figs. S8C and S9). This heterogeneity suggested that only a subset of REM sleep–active CM neurons densely projects to the dPFC.

To assess for a causal role for CM neurons in modulating the PV-PYR dPFC circuit during REM sleep, we optogenetically silenced the activity of CM neurons while conducting dual-color two-photon imaging of dPFC PYR and PV cell somas during REM sleep (Fig. 3, P and Q). PV-IRES-Cre animals were stereotactically injected with AAV1-CaMKII-SwichRCA in the CM and AAV-CaMKII-GCaMP6 (GCaMP6-expressing PYR somas) and AAV-Syn-Flex-NES-JRCaMP1b (RCaMP-expressing PV neurons) in the dPFC and were chronically implanted with EEG and EMG electrodes, an optical fiber placed above the CM area (with a 45° angle because of space constraints), and a cranial window above the dPFC (materials and methods, fig. S8D, and Fig. 3P). Optogenetic silencing of SwichRCA-expressing CM neurons induced a reliable and significant increase of PYR soma activity (Fig. 3, R to T), concomitant to a significant reduction of PV neuron activities during REM sleep, as compared with the control groups (Fig. 3, R, U, and V). Notably, the activity of both PYR and PV neurons, upon optogenetic silencing of CM neurons during REM sleep, was similar to the levels observed during wakefulness (Fig. 3, T and V).

VIP-mediated disinhibition of PYR dendrites

To test whether a causal link exists between REM sleep–specific activity enhancement of dPFC VIP neurons and PYR dendrites, VIP-IRES-Cre animals were stereotactically injected with AAV1-EF1-DIO-SwichRCA and AAV-CaMKII-GCaMP6 (diluted to obtain sparse labeling; see materials and methods) in the dPFC and were chronically implanted with EEG and EMG electrodes and a cranial window (Fig. 4, A and B). Simultaneous two-photon dendritic imaging and REM sleep–selective optical silencing of local VIP neurons was performed in superficial layers of the dPFC (Fig. 4C and materials and methods). Optogenetic silencing of SwichRCA-expressing VIP neurons induced a significant reduction of the frequency of both confined and spreading calcium events (Fig. 4, D and E, respectively), revealing a causal role for dPFC VIP neurons in the REM sleep–specific disinhibition of PYR dendrites.

Somatodendritic decoupling supports REM sleep–specific memory consolidation

Next, we assessed whether REM sleep somatodendritic decoupling is required for synaptic plasticity and the discrimination of dangerous versus safe environments using REM sleep–specific optogenetic silencing of PV and VIP neurons during an auditory fear-conditioning learning task (*11*, *30*, *31*). To optogenetically silence PV (i.e., somatic disinhibition) or VIP (i.e., dendritic inhibition) neurons, we genetically targeted the expression of archaerhodopsin (ArchT) to PV and VIP neurons by stereotactic injection of AAV2-EF1α-ArchT-eYFP or YFP tag (AAV2-EF1α-eYFP) as controls in the dPFC from PV-IRES-Cre and VIP-IRES-Cre mice, respectively (Fig. 5A). SST neurons were not included in this experiment because they were found to be inactive during REM sleep (Fig. 2, L to P).

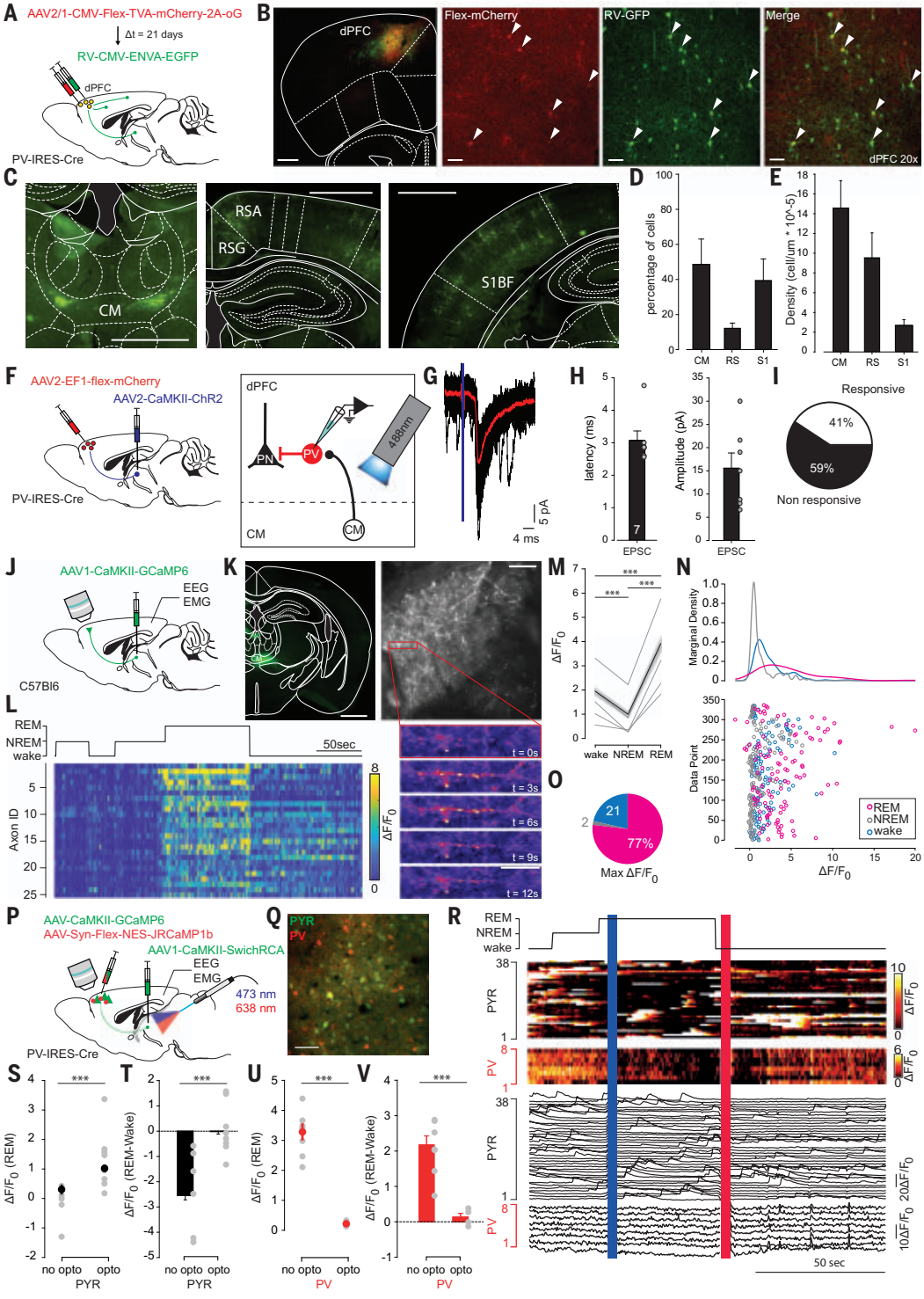


Fig. 3. Thalamic control of REM sleep–specific PV-mediated inhibition of PYR somas. (**A**) Schematic of the experimental procedure. (**B**) Viral injection site of flex-mCherry and RV-GFP in the dPFC. Scale bars, 500 μm (leftmost panel) and 50 μm (right three panels). (**C**) Expression profiles of the major brain regions projecting to dPFC PV neurons (CM, RSA and RSG, and S1BF). Scale bars, 1 mm. (**D**) Total distribution of identified neurons ± SEM (n = 3 mice) in the CM, RS, and S1. (**E**) Density of identified neurons ± SEM in the CM, RS, and S1. (**F**) Schematic of the experimental strategy. (**G**) Example trace of light-evoked EPSC in a PV neuron. Black traces indicate single evoked sweeps, the red trace indicates the sweeps' mean, and the blue vertical line indicates optical stimulation. (**H**) (Left) EPSC mean ± SEM latency from light stimulation (n = 7 neurons; m = 3 mice). (Right) Mean ± SEM amplitude of evoked EPSCs. (**I**) Percentages of cells responsive to light-evoked stimulation. (**J**) Schematic of the experimental procedure. (**K**) (Left) Expression profile of GCaMP6 in the CM. Scale bar, 1 mm. (Right) GCaMP6 CM axons. Scale bar, 30 μm. (Bottom) Time-lapse activity profile of a CM. Scale bar, 10 μm. (**L**) Hypnogram and color-coded $\Delta F/F_0$ activity profiles of CM axons. (**M**) Mean $\Delta F/F_0$ ± SEM (black line) of CM axons (n = 112 axons; m = 5 mice) during wake, NREM, and REM sleep. ***P < 0.001; one-way RM ANOVA with Tukey post hoc test. (**N**) Line (top) and scatter (bottom) plots showing the distribution of CM axons $\Delta F/F_0$ values during wake, NREM, and REM sleep. (**O**) Percentages of CM axons with maximal values of $\Delta F/F_0$ in wake (light blue), NREM (gray), and REM (fuchsia) sleep. (**P**) Schematic of the experimental procedure. (**Q**) GCaMP6 PYR neurons and RCaMP PV neurons. Scale bar, 30 μm. (**R**) Hypnogram and color-coded $\Delta F/F_0$ activity profiles of dPFC PYR and PV somas. Blue and red vertical lines indicate 473-nm and 638-nm optical stimulations, respectively. (**S**) REM sleep $\Delta F/F_0$ ± SEM of a group of PYR somas (n = 475 neurons; m = 8 mice) upon REM sleep–specific CM silencing (opto) versus a control group (no opto) (n = 238 neurons; m = 6 mice). ***P < 0.001; Student's t test. (**T**) PYR $\Delta F/F_0$ change ± SEM (difference between REM sleep and wake) in opto versus no opto. ***P < 0.001; Student's t test. (**U**) REM sleep $\Delta F/F_0$ ± SEM of PV neurons (n = 92 neurons; m = 4 mice) upon REM sleep–specific CM silencing (opto) versus control (no opto) (n = 137 neurons; m = 7 mice). ***P < 0.001; Student's t test. (**V**) PV $\Delta F/F_0$ change ± SEM (difference between REM sleep and wake) in opto versus no opto. ***P < 0.001; Student's t test.

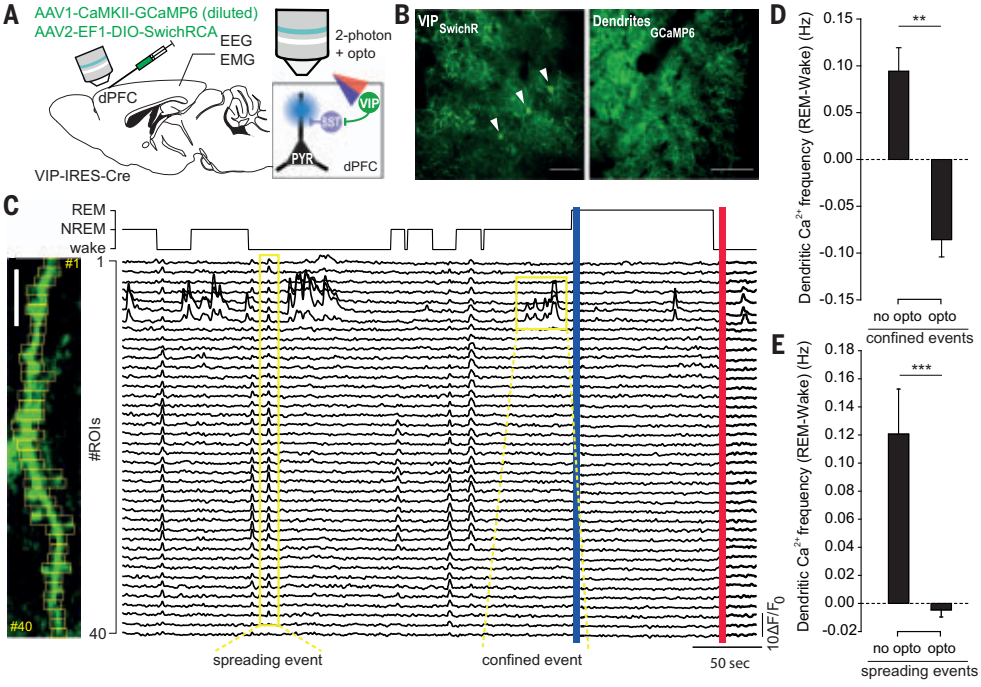


Fig. 4. VIP-mediated disinhibition of PYR dendrites. (**A**) Schematic of the experimental procedure. (**B**) GCaMP6 PYR dendrites (right) and SwichRCA VIP neurons (left). Scale bar, 30 μm. (**C**) Hypnogram and $\Delta F/F_0$ traces of ROIs obtained from the segmentation of a single dendrite (represented on the left). Scale bar, 10 μm. (**D**) Dendritic confined event frequency change ± SEM (difference between REM sleep and wake) in no opto (n = 308 dendrites; m = 5 mice) versus opto (n = 51 dendrites; m = 5 mice). $^{**}P < 0.01$; Student's t test. Raw frequency values during REM sleep: no opto = 0.5405 ± 0.03 Hz; opto = 0.0211 ± 0.009 Hz. (**E**) Dendritic spreading event frequency change ± SEM (difference between REM sleep and wake) in no opto versus opto. $^{***}P < 0.001$; Student's t test. Raw frequency values during REM sleep: no opto = 0.2012 ± 0.0326 Hz; opto = 0.0066 ± 0.004 Hz.

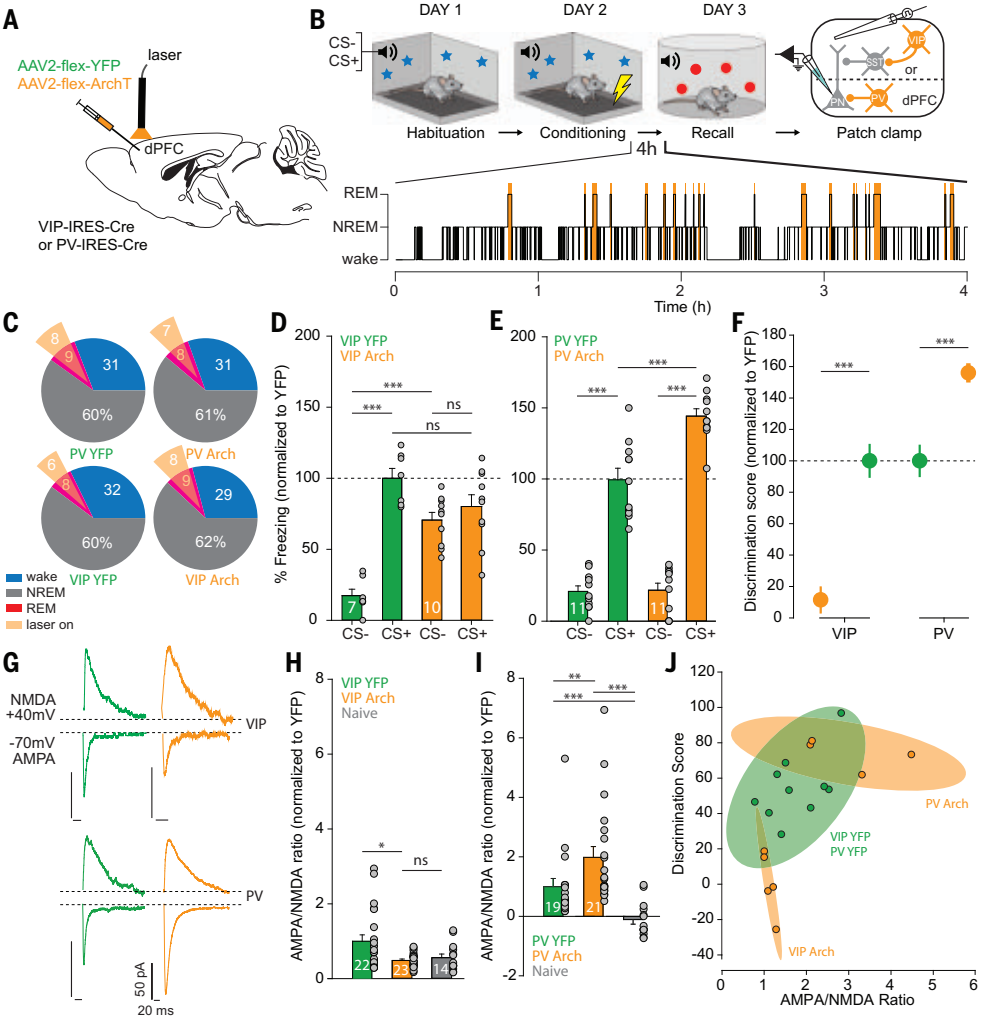


Fig. 5. VIP- and PV neuron–dependent tuning of REM sleep–specific memory consolidation. (**A**) Schematic of the experimental procedure. (**B**) Experimental timeline. Electrical stimulation is indicated in the diagram. (**C**) Sleep-wake state total duration during optogenetic manipulation in PV YFP (m = 11 mice), PV Arch (m = 11 mice), VIP YFP (m = 7 mice), and VIP Arch (m = 10 mice). (**D**) Freezing percentage (normalized to YFP) ± SEM during CS^- and CS^+ in VIP YFP (green) versus VIP Arch (orange). $^{***}P < 0.001$; ns, not significant; one-way ANOVA with Tukey post hoc test. (**E**) Freezing percentage (normalized to YFP) ± SEM during CS^- and CS^+ in PV YFP (green) versus PV Arch (orange). $^{***}P < 0.001$; ns, not significant; one-way ANOVA with Tukey post hoc test. (**F**) Discrimination score (normalized to YFP) ± SEM in VIP (left) and PV (right) mice (orange = Arch; green = YFP). $^{***}P < 0.001$; Student's t test. (**G**) AMPA/NMDA ratio traces. (**H**) Mean AMPA/NMDA ratios ± SEM (normalized to YFP) of VIP YFP (green) (n = 22 cells; m = 5 mice), VIP Arch (orange) (n = 23 cells; m = 5 mice), and naïve (gray, no behavior) (n = 14 cells; m = 3 mice). $^{*}P < 0.05$; ns, not significant; one-way ANOVA with Tukey post hoc test. (**I**) Mean AMPA/NMDA ratios ± SEM (normalized to YFP) of PV YFP (green) (n = 19 cells; m = 5 mice), PV Arch (orange) (n = 21 cells; m = 4 mice), and naïve (gray; normalized to YFP; raw value, 1.2832 ± 0.22). $^{**}P < 0.01$; $^{***}P < 0.001$; one-way ANOVA with Tukey post hoc test. (**J**) Correlation between discrimination score and AMPA/NMDA ratios. Estimations of 95% confidence ellipses are represented.

Subsequently, VIP ArchT and PV ArchT mice, and their respective YFP controls, were subjected to a fear-conditioning discriminative task, during which mice progressively learned to discriminate a negative (safety, CS^-) from a positive (danger, CS^+) conditional auditory stimuli (CS) (materials and methods). Mice showed no preference for either of the two conditional auditory stimuli (CS^+ and CS^-) during initial habituation (day 1) (fig. S10). After conditioning (day 2), all mice showed a significant increase of freezing in response to the CS^+ (day 1 versus day 3) (fig. S10). Immediately after the behavioral conditioning session (day 2), prefrontal ArchT-expressing VIP or PV neurons, and their respective YFP controls, were optogenetically silenced bilaterally selectively during every REM sleep episode using online detection of EEG and EMG signals over a 4-hour time window (532-nm continuous light) (Fig. 5, A and B). REM sleep–specific silencing of VIP or PV neurons did not perturb sleep-wake cycle architecture (Fig. 5C) or the probability of state transitions or sleep episode duration (fig. S11). The next day (day 3), mice were placed in a different context and exposed to CS^+ and CS^- without any aversive stimuli to retest for discriminative memory retention (Fig. 5B and fig. S10). In contrast to the YFP control mice, we observed that REM sleep–specific optogenetic silencing of VIP Arch mice lost the ability to discriminate CS^- from CS^+ (Fig. 5D) and showed a generalized freezing behavior (Fig. 5F), whereas optogenetic silencing of PV Arch mice led to a significant increase of CS^--versus-CS^+ discrimination (Fig. 5, E and F, and see fig. S10 for raw values). Accordingly, fear extinction was lower in PV Arch compared with control mice (fig. S12).

Finally, we tested whether optogenetic silencing of ArchT-expressing PV or VIP neurons affects synaptic plasticity within local prefrontal circuits using ex vivo acute patch-clamp recordings after memory recall and REM sleep–specific optogenetic silencing (day 3) (Fig. 5, B and G). Consistent with the behavioral results, dPFC PYR neurons displayed a significant decrease of AMPA/NMDA (*N*-methyl-D-aspartate) ratio after REM sleep–silencing of ArchT-expressing VIP neurons, which suggests a suppression of local synaptic plasticity (Fig. 5H). By contrast, they exhibited a significant enhancement of AMPA/NMDA ratio after silencing of ArchT-expressing PV neurons as compared with YFP control, suggesting an increase of local synaptic plasticity (Fig. 5I). Correlation analysis between AMPA/NMDA ratio and discrimination score further showed a segregation between VIP Arch mice (low AMPA/NMDA ratio and discrimination values), PV Arch mice (high AMPA/NMDA ratio and discrimination values), and their respective PV-YFP and VIP-YFP control groups (Fig. 5J).

Discussion

Collectively, these results demonstrate that the E/I balance within dPFC microcircuits is altered specifically during REM sleep through PV-mediated inhibition of PYR somas concomitant to a VIP-dependent disinhibition of PYR dendritic activity, presumably because of the inhibition of SST neurons, as described ex vivo (*32*). Furthermore, the activity of PV neurons during REM sleep is controlled by inputs from a subset of mediodorsal thalamic neurons (*33*), as described in somatosensory (*32*, *34*) and associative (*33*) thalamocortical circuits. This circuit model and the inputs to prefrontal interneurons may differ in deeper cortical layers and the rostrocaudal location of the input cell bodies, respectively, according to the neuromodulation associated with sleep-wake states and their transitions.

A consequence of this somatodendritic decoupling during REM sleep is the consolidation of emotional memories through a (paradoxical) increase of excitability of PYR dendrites concomitant with a tonic inhibition of PYR somas, associated with a decrease of neuronal excitability. Although dendritic disinhibition during REM sleep provides a window to consolidate information acquired during wakefulness, somatic inhibition may preclude a top-down reinforcement of physiological and behavioral responses to fear in downstream pathways.

A lack of somatic inhibition may interfere with local dendritic computation through back-propagating activity (*35*, *36*) and facilitate nonlinear integration of synaptic inputs during REM sleep. This may result in overconsolidation of emotional memories observed in post-traumatic stress disorders and other affective psychiatric and mood disorders often associated with REM sleep disturbances (*37*, *38*).

These findings are consistent with a simultaneous upscaling and downscaling of synaptic plasticity at the PYR dendrites and cell bodies, respectively. Yet, whether similar NREM sleep–dependent mechanisms support memory consolidation and synaptic plasticity remains to be investigated (*39*, *40*). Somatodendritic decoupling during REM sleep provides a mechanism that allows flexibility or adaptation to environmental changes and optimizes the discrimination of danger versus safety and physiological responses to behavioral stressors.

REFERENCES AND NOTES

1. I. Léna *et al.*, *J. Neurosci. Res.* **81**, 891–899 (2005).
2. L. Dahan *et al.*, *Neuropsychopharmacology* **32**, 1232–1241 (2007).
3. K. J. Maloney, L. Mainville, B. E. Jones, *Eur. J. Neurosci.* **15**, 774–778 (2002).
4. P. Maquet *et al.*, *Nature* **383**, 163–166 (1996).
5. A. R. Braun *et al.*, *Brain* **120**, 1173–1197 (1997).
6. E. A. Nofzinger, M. A. Mintun, M. Wiseman, D. J. Kupfer, R. Y. Moore, *Brain Res.* **770**, 192–201 (1997).
7. S. Ribeiro *et al.*, *J. Neurosci.* **22**, 10914–10923 (2002).
8. P. Ravassard *et al.*, *Sleep* **32**, 227–240 (2009).
9. G. R. Poe, D. A. Nitz, B. L. McNaughton, C. A. Barnes, *Brain Res.* **855**, 176–180 (2000).
10. J. Seibt *et al.*, *Nat. Commun.* **8**, 684 (2017).
11. Y. Zhou *et al.*, *Nat. Commun.* **11**, 4819 (2020).
12. R. Boyce, S. D. Glasgow, S. Williams, A. Adamantidis, *Science* **352**, 812–816 (2016).
13. C. Pavlides, J. Winson, *J. Neurosci.* **9**, 2907–2918 (1989).
14. W. Li, L. Ma, G. Yang, W. B. Gan, *Nat. Neurosci.* **20**, 427–437 (2017).
15. G. R. Poe, *J. Neurosci.* **37**, 464–473 (2017).
16. S. Izawa *et al.*, *Science* **365**, 1308–1313 (2019).
17. F. Crick, G. Mitchison, *Nature* **304**, 111–114 (1983).
18. R. Stickgold, M. P. Walker, *Nat. Neurosci.* **16**, 139–145 (2013).
19. M. A. Kheirbek *et al.*, *Neuron* **77**, 955–968 (2013).
20. M. D. Morrissey, K. Takehara-Nishiuchi, *Neurobiol. Learn. Mem.* **115**, 95–107 (2014).
21. R. Maciel *et al.*, *Biochem. Pharmacol.* **191**, 114514 (2021).
22. L. Renouard *et al.*, *Sci. Adv.* **1**, e1400177 (2015).
23. N. Niethard *et al.*, *Curr. Biol.* **26**, 2739–2749 (2016).
24. J. P. Shaffery, C. M. Sinton, G. Bissette, H. P. Roffwarg, G. A. Marks, *Neuroscience* **110**, 431–443 (2002).
25. L. I. Schmitt *et al.*, *Nature* **545**, 219–223 (2017).
26. J. S. Stevens *et al.*, *J. Psychiatr. Res.* **47**, 1469–1478 (2013).
27. B. R. Ferguson, W. J. Gao, *Front. Neural Circuits* **12**, 37 (2018).
28. H. J. Pi *et al.*, *Nature* **503**, 521–524 (2013).
29. F. Osakada, E. M. Callaway, *Nat. Protoc.* **8**, 1583–1601 (2013).
30. C. S. W. Lai, T. F. Franke, W.-B. Gan, *Nature* **483**, 87–91 (2012).
31. M. Aime *et al.*, *eLife* **9**, e62594 (2020).
32. L. E. Williams, A. Holtmaat, *Neuron* **101**, 91–102.e4 (2019).
33. A. Mukherjee, N. H. Lam, R. D. Wimmer, M. M. Halassa, *Nature* **600**, 100–104 (2021).
34. B. S. Sermet *et al.*, *eLife* **8**, e52665 (2019).
35. S. L. Smith, I. T. Smith, T. Branco, M. Häusser, *Nature* **503**, 115–120 (2013).
36. M. Larkum, *Trends Neurosci.* **36**, 141–151 (2013).
37. E. F. Pace-Schott, A. Germain, M. R. Milad, *Biol. Mood Anxiety Disord.* **5**, 3 (2015).
38. A. L. A. Murkar, J. De Koninck, *Sleep Med. Rev.* **41**, 173–184 (2018).
39. B. Rasch, J. Born, *Physiol. Rev.* **93**, 681–766 (2013).
40. G. Tononi, C. Cirelli, *Eur. J. Neurosci.* **51**, 413–421 (2020).

ACKNOWLEDGMENTS

We thank all the Tidis Laboratory members, M. Schmidt, C. Gutierrez Herrera, M. Baud, and C. Bassetti for their insightful discussions and comments on previous versions of the manuscript; K. Deisseroth, A. Holtmaat, and FMI Basel Vector Core for providing viral tools for optogenetics, dual-color two-photon imaging experiments, and retrograde monosynaptic tracing, respectively; S. Sachidhanandam for critical evaluation of the manuscript; and the S. Saxena Laboratory for providing support for confocal imaging. **Funding:** This work was supported by the Inselspital University Hospital Bern, the European Research Council (CoG-725850 to A.A.), the Swiss National Science Foundation (A.A.), the Synapsis Foundation (A.A.), and the University of Bern (A.A.). **Author contributions:** M.A., M.B., and T.C. performed the experiments. M.A., N.C., M.B., and T.R. analyzed the data. M.A. and A.A. conceived the study and wrote the manuscript with the help of all authors. A.A. supervised the research. A.S. and T.F. provided technical assistance with the design and fabrication of aberration-corrected microendoscopes. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All presented data and analysis scripts, including mat-files and Matlab scripts and functions, are available on https://github.com/ZENLabCode.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk2734
Materials and Methods
Figs. S1 to S13
References (*41*–*43*)
MDAR Reproducibility Checklist
Movie S1

Submitted 5 July 2021; resubmitted 24 January 2022
Accepted 4 April 2022
10.1126/science.abk2734

FERROELECTRICS

Highly enhanced ferroelectricity in HfO_2-based ferroelectric thin film by light ion bombardment

Seunghun Kang[1]†, Woo-Sung Jang[2]†, Anna N. Morozovska[3]†, Owoong Kwon[1], Yeongrok Jin[4], Young-Hoon Kim[2], Hagyoul Bae[5], Chenxi Wang[1], Sang-Hyeok Yang[2], Alex Belianinov[6,7], Steven Randolph[6], Eugene A. Eliseev[8], Liam Collins[6], Yeehyun Park[1], Sanghyun Jo[5], Min-Hyoung Jung[2], Kyoung-June Go[9], Hae Won Cho[1], Si-Young Choi[9], Jae Hyuck Jang[10], Sunkook Kim[1], Hu Young Jeong[11], Jaekwang Lee[4], Olga S. Ovchinnikova[12], Jinseong Heo[5]*, Sergei V. Kalinin[6,13]*, Young-Min Kim[2]*, Yunseok Kim[1]*

Continuous advancement in nonvolatile and morphotropic beyond-Moore electronic devices requires integration of ferroelectric and semiconductor materials. The emergence of hafnium oxide (HfO_2)–based ferroelectrics that are compatible with atomic-layer deposition has opened interesting and promising avenues of research. However, the origins of ferroelectricity and pathways to controlling it in HfO_2 are still mysterious. We demonstrate that local helium (He) implantation can activate ferroelectricity in these materials. The possible competing mechanisms, including He ion–induced molar volume changes, vacancy redistribution, vacancy generation, and activation of vacancy mobility, are analyzed. These findings both reveal the origins of ferroelectricity in this system and open pathways for nanoengineered binary ferroelectrics.

Progress toward ultradense oxide electronics and applications, including nonvolatile memories and beyond-Moore devices, necessitates the introduction of improved device paradigms and the development of strategies that utilize the wealth of functionalities of complex materials at extremely reduced dimensions. These include a wide range of emergent properties, such as magnetoelectricity, ferroelectricity, flexoelectricity, and electrochromicity, that result from the intrinsic competition (or cooperation) of charge, spin orbital, and lattice degrees of freedom (*1*). The switching, logical, and memory devices require a special role for ferroelectricity, which refers to a material property of spontaneous electrical polarization that is reversible when an electric field is applied. Ferroelectricity has been studied for the application of next-generation electronic devices, such as ferroelectric field-effect transistors (*2*, *3*) and nonvolatile memories (*4*, *5*), for over 40 years owing to their desirable physical properties, including a fast-switching speed and long-term stability (*6*). However, despite these superior properties, ferroelectrics have practical limitations owing to scaling effects (*7*, *8*) and a low compatibility with complementary metal-oxide semiconductor (CMOS) processes (*9*). In particular, the ferroelectricity of conventional perovskite ferroelectrics substantially drops at thicknesses below 50 nm, limiting their application in ultralarge-scale integrated electronics (*7*). Second, integration between classical ferroelectrics and Si and other semiconductors is limited by intrinsically high oxygen chemical potential in the former, leading to formation of parasitic SiO_x layers and degradation during classical processing. Hence, despite much effort, the goal of ferroelectric-semiconductor integration has remained elusive.

The discovery of HfO_2-based ferroelectrics with a fluorite structure in 2011 (*10*), e.g., $Hf_{0.5}Zr_{0.5}O_2$ (HZO), broke the traditional paradigm and offers the promise of nonvolatile and morphotropic beyond-Moore electronic devices (*10*–*13*). Unlike conventional ferroelectrics, HfO_2-based ferroelectrics maintain robust ferroelectricity (*14*), which would be viable even down to a fundamental barrier restricted by a unit cell dimension, and are highly compatible with the existing CMOS processes (*15*). However, the nature of ferroelectricity in these materials remains poorly understood owing to the complexity that arises from undefined structural and chemical factors. This poor understanding has in turn hindered the rapid development of facile controllability and large-scale uniformity. From a phenomenological perspective, HfO_2-based ferroelectrics demand a "wake-up" process to increase ferroelectricity by electric field cycling (*16*–*20*), suggesting the nonconventional nature of ferroelectric responses. Furthermore, the deposition parameters, such as doping, oxygen partial pressure and annealing conditions (*21*–*24*), and engineering perspective after surface modifications such as plasma treatment (*25*, *26*), were found to have a strong effect on the resulting ferroelectricity.

Ferroelectricity in HfO_2-based ferroelectrics can originate from multiple mechanisms, such as intrinsic ferroelectricity (*27*), electret-like responses associated with charge trapping at internal interfaces and intrinsic electrostriction (*28*). Moreover, recent studies have illustrated the important role of oxygen vacancies (*29*). This has gained particular attention in HfO_2-based ferroelectrics because the vacancies stabilize the metastable ferroelectric orthorhombic ($Pca2_1$) phase over the entire polycrystalline HZO thin film, resulting in an enhancement of ferroelectricity (*23*, *30*). Oxygen vacancy redistribution has been suggested as a possible mechanism for enhancement of the ferroelectricity driven by the wake-up effect (*16*–*19*). Despite these findings, controlling the ferroelectric response in HfO_2-based ferroelectrics has remained elusive.

We propose a strategy for highly enhanced and precisely controlled ferroelectricity in HfO_2-based thin films by point defect engineering through light-ion bombardment. We also report possible competing mechanisms that include molar volume changes, vacancy distribution, number of vacancies, and vacancy mobility. We provide evidence for vacancy generation as the dominant mechanism behind the emergence of ferroelectric behavior in these materials.

To quantitatively evaluate the effective enhancement of ferroelectricity after ion beam treatment, we explore the remnant polarization states and the hysteretic behavior via piezoresponse force microscopy (PFM) combined with various advanced techniques, such as the band excitation (BE) method and laser Doppler vibrometer (LDV). To understand the origin of the enhanced ferroelectricity, we examined the atomic structure–chemistry relationship in the polycrystalline HZO thin films before and after the light-ion treatment based on scanning transmission electron microscopy (STEM) combined with energy dispersive x-ray spectroscopy (EDX) and electron energy loss spectroscopy (EELS). We revealed that oxygen vacancies are homogeneously distributed over the optimally treated HZO thin film without the intervention of structural defects, which eventually leads to a global phase transition

[1]School of Advanced Materials Science and Engineering, Sungkyunkwan University (SKKU), Suwon 16419, Republic of Korea. [2]Department of Energy Science, Sungkyunkwan University (SKKU), Suwon 16419, Republic of Korea. [3]Institute of Physics, National Academy of Sciences of Ukraine, 46, Prospekt. Nauky, 03028 Kyiv, Ukraine. [4]Department of Physics, Pusan National University, Busan 46241, Republic of Korea. [5]Samsung Advanced Institute of Technology, Suwon 16678, Republic of Korea. [6]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [7]Sandia National Laboratories, Albuquerque, NM 87123, USA. [8]Institute for Problems of Materials Science, National Academy of Sciences of Ukraine, Krjijanovskogo 3, 03142 Kyiv, Ukraine. [9]Department of Materials Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea. [10]Center for Scientific Instrumentation, Korea Basic Science Institute (KBSI), Daejeon 34133, Republic of Korea. [11]Graduate School of Semiconductor Materials and Devices Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea. [12]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [13]Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN 37920, USA.
***Corresponding author. Email: jinseong.heo@samsung.com (J.H.); sergei2@utk.edu (S.V.K.); youngmk@skku.edu (Y.-M.K.); yunseokkim@skku.edu (Y.K.)**
†These authors contributed equally to this work.

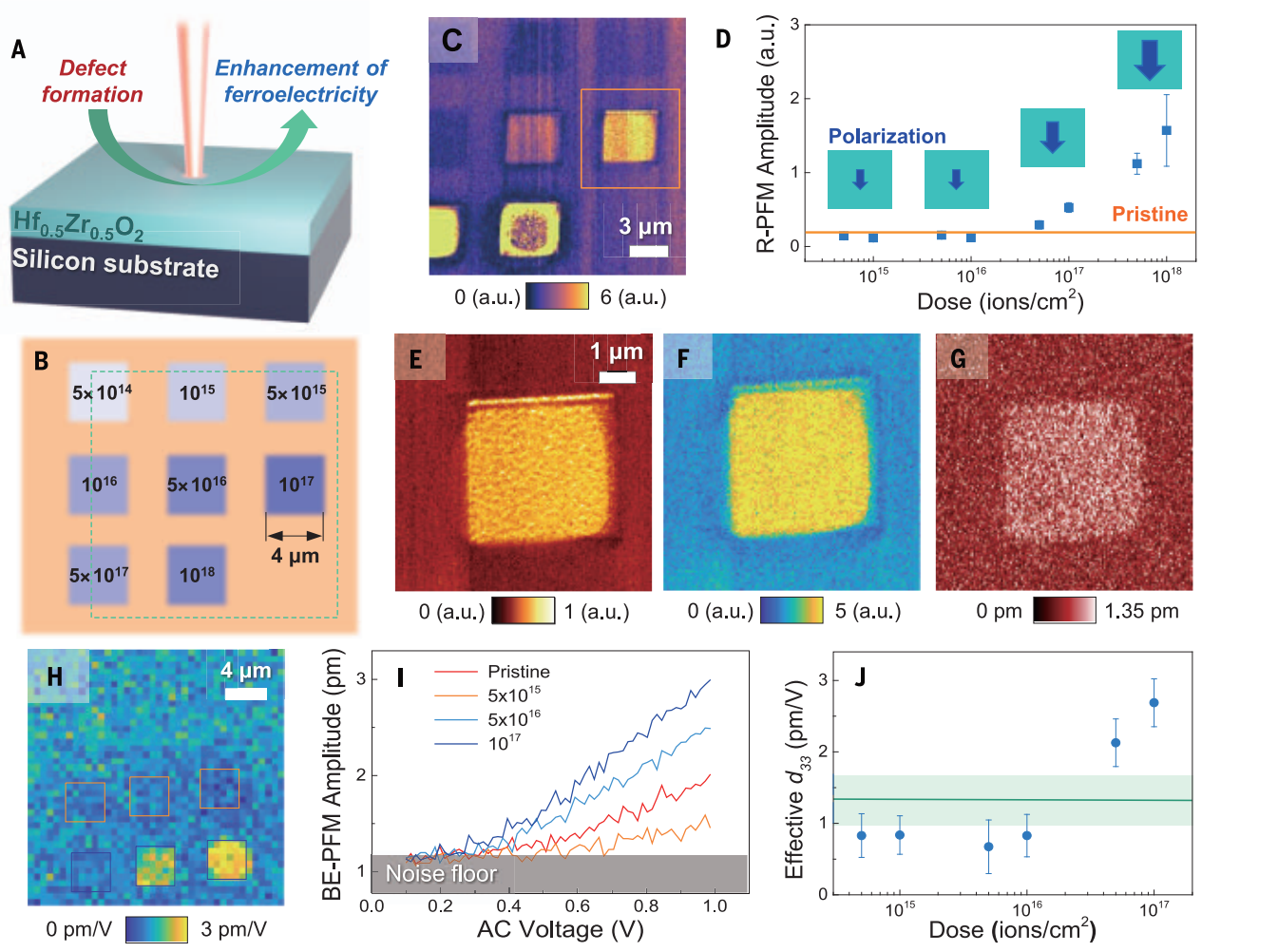


Fig. 1. Design of experiments and spontaneous polarization properties. (**A**) Schematic representation of ion bombardment process. (**B**) Scheme for dose of He ion bombardment; inserted texts indicate the He ion doses (ions/cm^2). (**C**) R-PFM amplitude image of HZO thin film after ion bombardment; green-dotted box in (B) indicates the measurement position of (C). (**D**) R-PFM amplitudes as a function of He ion dose extracted from (C) except for the central part of 10^{18} ions/cm^2. (**E**) R-PFM, (**F**) BE-PFM, and (**G**) LDV-PFM amplitude images for 10^{17} ions/cm^2 corresponding to the orange dotted box in (C). The PFM image is rotated 90° counterclockwise to align according to (B) and (C). (**H**) Spatial map of effective d_{33} after ion bombardment in (B); the orange and blue dashed boxes serve as visual guides corresponding to the positions with the He ion dose of 5×10^{14} to 5×10^{15} and 10^{16} to 10^{17} ions/cm^2, respectively. (**I**) BE-PFM amplitude versus V_{ac} depending on the He ion dose; BE-PFM amplitude was calibrated via the force–distance curve–based static calibration method (*55*). (**J**) Effective d_{33} as a function of He ion dose; the green line and area indicate the mean ± SD deviation of the effective d_{33} in the pristine region, respectively. The effective d_{33} results of (H) and (J) were extracted using least-squares linear fitting, except for the noise floor region.

to the ferroelectric orthorhombic phase whereas the nontreated or unoptimized thin films present a mixture of nonferroelectric and ferroelectric phases. This result implies a strong coupling of the oxygen vacancy and the resulting ferroic behavior of the HZO thin film. Thermodynamical interpretation can elucidate the phenomenological observation as possible competing mechanisms, including molar volume changes, vacancy distribution, number of vacancies, and vacancy mobility. The method that we propose to modulate the density of the point defects could be a breakthrough for the practical application of ferroelectrics to next-generation memory devices via enhanced ferroelectricity.

Piezoresponse-dependent light ion bombardment

We chose He ion bombardment as a strategy to explore the emergence of ferroelectricity, where He ions can become trapped by the lattice, leading to an increase in the molar volume, as well as generating oxygen vacancies. We performed ion bombardment on the HZO thin film as a function of the He ion dose using focused-beam He ion microscopy (Fig. 1A). To observe the change in the electric polarization as a function of defect density, we bombarded the thin film with various He ion doses (Fig. 1B). We determined the changes in the local polarization associated with the He ion dose using resonance PFM (R-PFM), which was measured with an AC voltage of a single frequency near the contact resonance (Fig. 1, C and D) (*31*).

The regions with He ion doses below 10^{16} ions/cm^2 do not demonstrate noticeable changes in the R-PFM amplitude, which represents the magnitude of the piezoresponse signal that in turn is related to polarization. Instead, these regions present a slight decrease in the piezoresponse. This reduction may result from the defect formation by ion bombardment. However, it may be also associated with carbon contamination at the surface because carbon deposition can be easily induced by the surface adsorbates of hydrocarbon molecules during ion bombardment (*32*–*34*) and, furthermore,

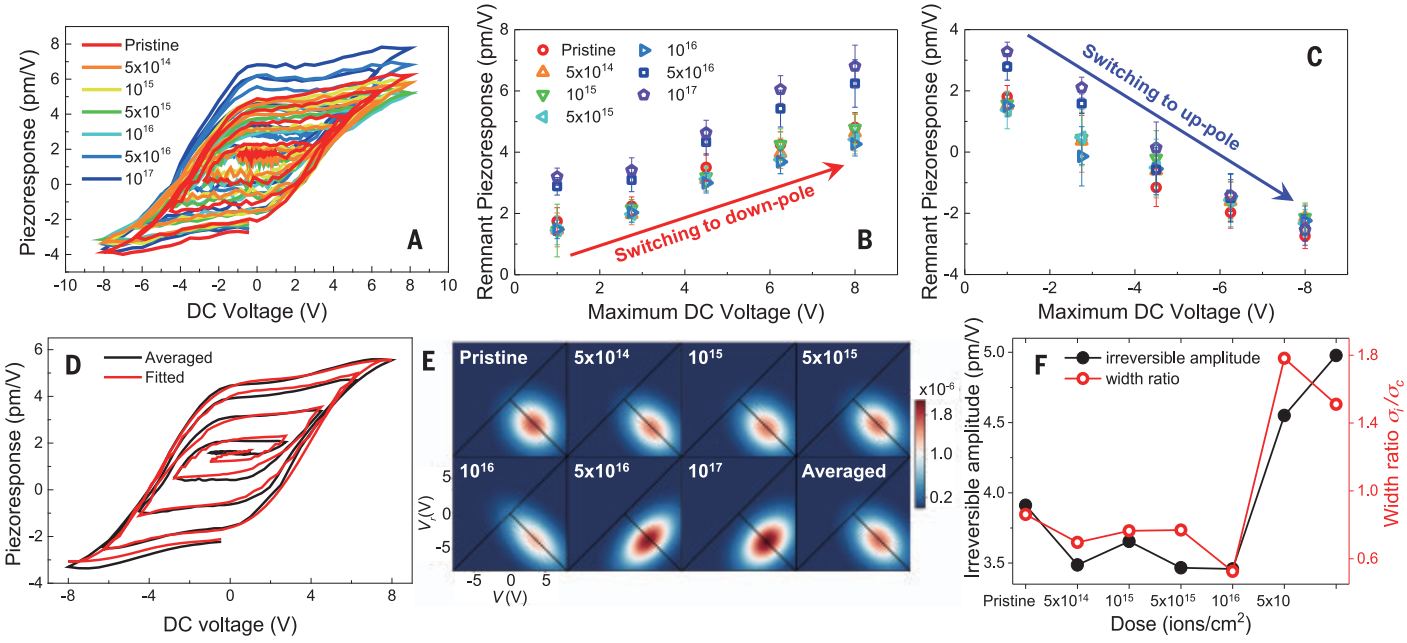


Fig. 2. Switchable polarization properties. (**A**) FORC-PFM hysteresis loop depending on the He ion dose. (**B**) Positive and (**C**) negative remnant piezoresponse versus $V_{dc,max}$ depending on the He ion dose. (**D**) Average piezoresponse hysteresis loop and the fitting result by the Preisach model. (**E**) Distribution of Preisach density depending on the He ion dose. (**F**) Dependence of the irreversible amplitude and the ratio of the interaction and coercivity widths extracted by the Preisach model on the He ion dose.

the atomic force microscopy (AFM) measurements can be sensitive to carbon contamination. However, the regions with He ion doses above 10^{16} ions/cm^2 show an increase in the R-PFM amplitude with an increase in the He ion dose. In particular, the R-PFM amplitude in the region with the He ion dose of 10^{17} ions/cm^2 increased by approximately two times compared with the pristine region, indicating substantial enhancement of ferroelectricity. Meanwhile, the central part of the region with the He ion dose of 10^{18} ions/cm^2 shows a rapid decrease in the R-PFM amplitude and the milled surface (fig. S1) because an excessive He ion dose can lead to a crystallographic collapse via material sputtering.

R-PFM can be affected by other contributions, such as a resonance shift or nonferroelectric contributions (*35*–*38*), so we measured BE-PFM and LDV-PFM (Fig. 1, F and G) in addition to R-PFM (Fig. 1E). Because BE-PFM can track the resonance frequency, it allows one to minimize the error from the resonance shift (*39*), which can be associated with carbon contamination. Similar to the R-PFM results, BE-PFM amplitude also indicates an increase in the piezoresponse after ion bombardment. In addition, we confirmed that there was little change of mechanical properties in the ion-bombarded HZO thin film through BE-PFM (fig. S2). Meanwhile, recent studies have reported that nonferroelectric factors, such as the electrostatic effect, can be attributed to the PFM signal (*35*, *36*). The difference in the surface potential according to the ion bombardment may affect the PFM signal due to electrostatic effects (*40*).

Therefore, we conducted LDV-PFM measurements in the same region (Fig. 1G) to evaluate the PFM signal, excluding the long-range electrostatic effect because LDV-PFM is less sensitive to the electrostatic effect (*40*, *41*). We observed a higher PFM amplitude in the region with a He ion dose of 10^{17} ions/cm^2. Furthermore, although the surface potential in the ion-bombarded region relaxed to the same value as that in the pristine region after a sufficient time elapsed, the R-PFM amplitude still exhibited a strong response in the ion-bombarded region (see fig. S3, 10^{17} ions/cm^2), indicating that the highly increased PFM signal originated primarily from the enhanced polarization by the ion bombardment. In addition, to verify the reproducibility of the thin films deposited by another method, we measured R-PFM images after ion bombardment on the HZO thin film deposited by the atomic layer deposition (fig. S4). We observed that light-ion bombardment continues to enhance the polarization in the HZO thin film, regardless of the deposition method used. Overall, the R-PFM results indicate that ion bombardment can be used to control the polarization by controlling the He ion dose.

We measured the effective d_{33} to explore the change in the polarization of HZO after ion bombardment, which may indicate the out-of-plane piezoelectric constant of spontaneous polarization, through an alternating-current voltage (V_{ac}) amplitude sweep in the HZO thin film. We determined the variation of the effective d_{33} obtained from the mapping result of the V_{ac} amplitude sweep, i.e., the BE-PFM amplitude according to V_{ac}, after ion bombardment (Fig. 1H). The regions that we measured correspond to the regions of 5×10^{14} to 10^{17} ions/cm^2 (Fig. 1, I and J). We excluded the regions of 5×10^{17} to 10^{18} ions/cm^2 because of topographic variations, including beam damage (fig. S1). The regions with He ion doses of 5×10^{16} to 10^{17} ions/cm^2 clearly have a higher effective d_{33} compared with that of the other regions (Fig. 1, H and I), although the difference according to the He ion dose was not observed owing to the equipment background noise below 0.3 V_{ac}. We averaged the effective d_{33} results for a quantitative comparison as a function of the He ion dose (Fig. 1J). The regions with He ion doses below 10^{16} ions/cm^2 show a slight decrease in the effective d_{33}. This decrease may be due to the defect formation by ion bombardment or the carbon contamination at the surface of the HZO thin film after ion bombardment (*32*–*34*). When the He ion dose is increased to 5×10^{16} ions/cm^2 or more, the effective d_{33} is substantially increased compared with that of the pristine region. In particular, the effective d_{33} value of the 10^{17} ions/cm^2 region increased by 99% compared with that of the pristine region. Therefore,

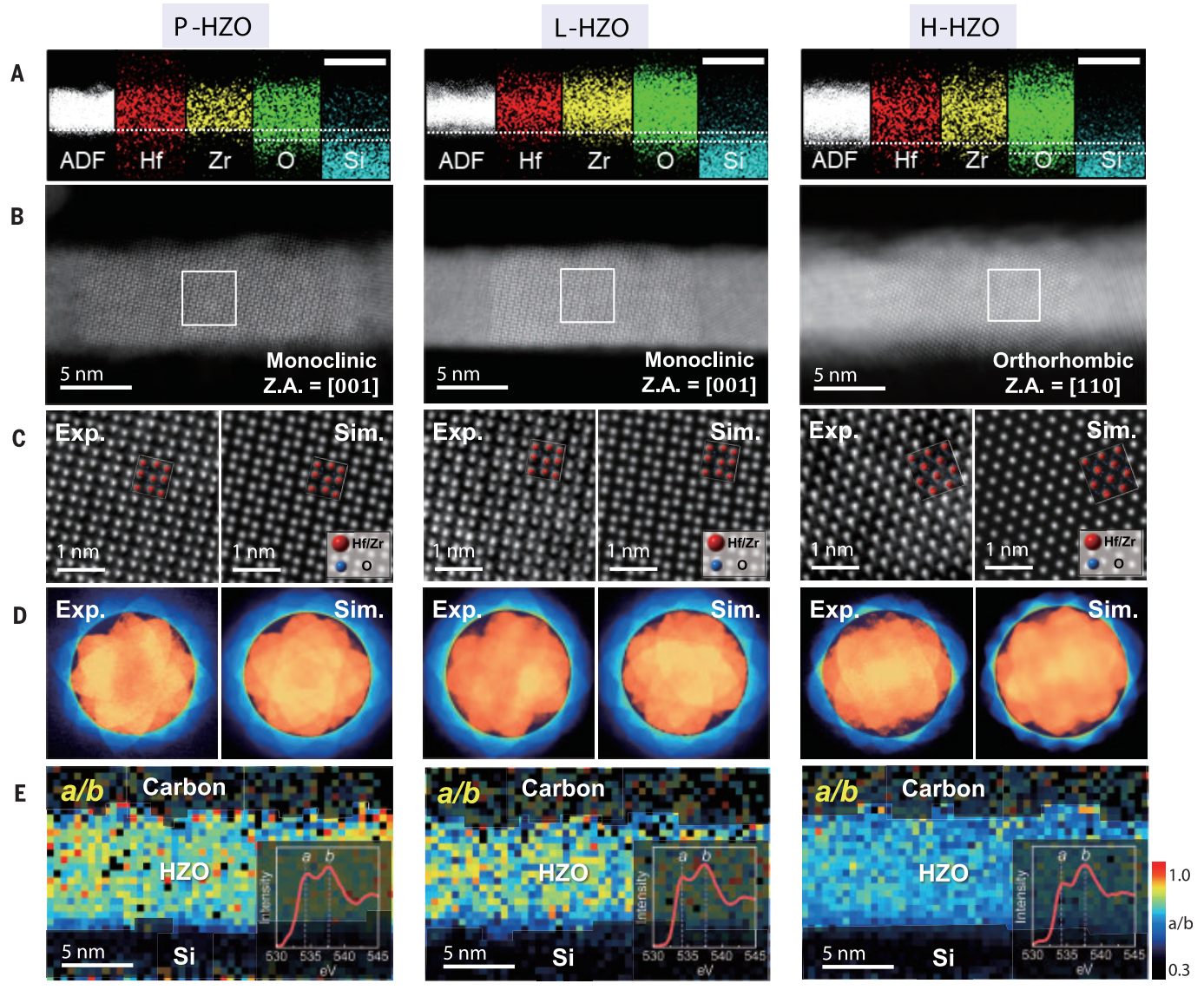


Fig. 3. Chemical and structural analysis of crystal phases and defects. (**A**) STEM-EDX chemical maps of the constituent elements Hf, Zr, O, and Si in the HZO samples with different He ion dose conditions: P-HZO, L-HZO, and H-HZO, respectively. For elemental mapping, characteristic x-ray peaks of Hf L (7.898 keV), Zr K (15.744 keV), O K (0.525 keV), and Si K (1.739 keV) were selected. The weak Si contrast shown in the HZO thin film is attributed to the peak overlap of Si K and Hf M (1.644 keV) peaks. Scale bars, 10 nm. (**B**) Representative HAADF images of the three HZO samples. One of the nanograins in the three samples was aligned to a crystal zone-axis for atomic resolution imaging. (**C**) (Left) Enlarged atomic structure images for the regions marked by white boxes in (B); (right) simulated HAADF images corresponding to the experimental images. The Hf and Zr lattice models used for the HAADF image simulations were overlaid on each image for comparison. The aligned nanograins in the three HZO samples were identified to be [001]-oriented monoclinic and [110]-oriented orthorhombic structures, respectively. (**D**) Comparison between the experimental and the simulated PACBED patterns obtained from each nanograin in the three HZO samples. The simulated PACBED patterns were calculated from the three identified atomic structures. (**E**) Maps of *a*/*b* ratio in O K EELS for the three HZO samples. The maps were reconstructed by NLLS fitting process for the obtained EELS SI datasets. Insets represent O K EELS spectra averaged over the HZO film regions, respectively. The outside regions shaded with dark gray in each HZO sample were not considered for the evaluation of the *a*/*b* ratio because those regions are unrelated to the HZO film. The color scale bar on the rightmost side indicates the range of the measured *a*/*b* ratio in O K edge.

the PFM results indicate that light-ion bombardment can enhance the polarization of HZO; furthermore, it can be used to locally control these enhancements. However, considering the applicability of ferroelectrics to next-generation electronic devices, the switching characteristics of the polarization rather than a spontaneous polarization are more important.

Influence of light-ion bombardment on polarization switching

To observe the changes in the polarization switching characteristics owing to He ion bombardment, we performed PFM hysteresis loops based on the first-order reversal-curve (FORC) method (*42*). We measured the polarization switching behavior depending on the maximum direct-current voltages ($V_{dc,max}$) as a function of the He ion dose (Fig. 2A). Whereas the coercive voltages related to the polarization switching were in a similar range regardless of the ion bombardment and He ion dose, the piezoresponse indicating the intensity and direction of polarization was substantially increased at

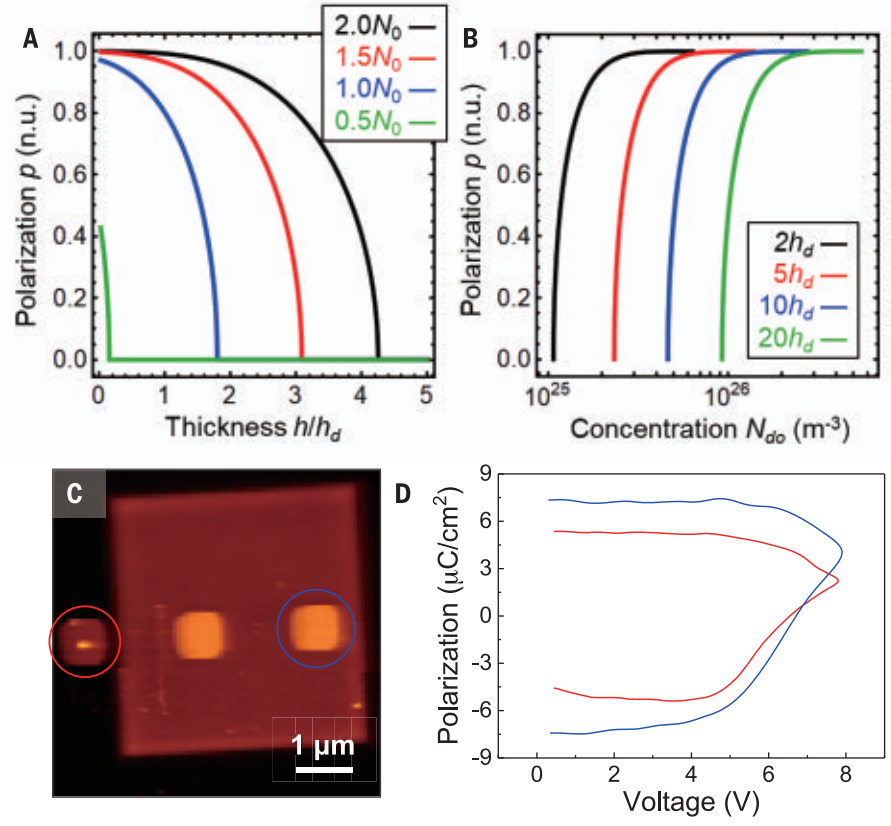


Fig. 4. Polarization charge after ion bombardment. Normalized polarization $P(E)$ on the electric field E calculated for (**A**) normalized spontaneous polarization p versus relative film thickness h/h_d calculated for different N_{do} = (2, 1.5, 1, 0.5)·N_0 (black, red, blue, and green curves, respectively) where $N_0 \cong 10^{25}$ m^{-3} and h_d is the decay factor of the defect concentration. (**B**) Normalized spontaneous polarization p versus vacancy concentration N_{do} calculated for different film thicknesses h = (2, 5, 10, 20) h_d (black, red, blue, and green curves, respectively). Parameters: T = 300 K, $\frac{J}{k_B} = -10$ K, $q_{33}W_{33}^{d}N_0 = -5 \cdot 10^6$ m J/C^2 corresponds to, e.g., $N_0 \cong 10^{25}$ m^{-3} at $W_{33}^{d} \cong -10^{-29}$ m^{-3} and $q_{33} = 5 \cdot 10^{10}$ m J/C^2. (**C**) Topographic image of HZO thin film with top electrodes. Red and blue circles indicate the points where P-V curve measurements by AFM-PUND were performed for the pristine and bombarded regions, respectively. The slightly larger height with rectangular shape indicates the region bombarded with 10^{17} ions/cm^2. (**D**) P-V curves of the pristine (red) and bombarded (blue) regions. In the case of the pristine region, P-V curve measurements were performed before He ion bombardment.

He ion doses above 10^{16} ions/cm^2. These results can be more clearly observed through the remnant piezoresponse (Fig. 2, B and C). The positive remnant piezoresponse increases more steeply in the He ion doses of 5×10^{16} and 10^{17} ions/cm^2 than that of other doses with an increasing $V_{dc,max}$ as the polarization direction shifts downward in the HZO thin film, except for the result at 1 $V_{dc,max}$, where switching does not occur (Fig. 2B). In the region with a He ion dose of 10^{17} ions/cm^2, the positive remnant piezoresponse at 8 $V_{dc,max}$ was increased by ~42% compared with that in the pristine region. The ion bombardment in the present case does not substantially affect leakage current (fig. S5). The piezoresponse was slightly decreased at He ion doses below 10^{16} ions/cm^2, indicating that the slight decrease in both R-PFM amplitude and remnant piezoresponse may have resulted from the defect formation by ion bombardment as well, rather than the carbon contamination.

The PFM hysteresis loops that we obtained by the FORC method were well-fitted by the generalized Preisach model (Fig. 2D) (*31*). Our analysis suggests that He ion bombardment yields broader distribution of coercive fields than the internal bias fields for the average result, suggesting the nonuniform nature of the material. However, the distributions of the Preisach density for the He ion dose of 5×10^{16} to 10^{17} ions/cm^2 are clearly distinguished from those for the other doses by the narrower distribution of the coercive voltages and higher Preisach density than the coercive fields (Fig. 2E), indicating more uniform and strengthened polarization switching responses. In addition, because of the increasing defect density, the interaction between charged defects and the spontaneous polarization is enhanced, leading to an extended distribution of internal bias voltages for the He ion dose of 5×10^{16} to 10^{17} ions/cm^2. Apparently, a threshold value exists for the He ion dose to induce strengthened polarization switching and strong interactions. The ratio of the internal bias voltage width to the coercive voltage width and the irreversible amplitude both present an abrupt increment between the He ion doses of 10^{16} and 5×10^{16} ions/cm^2 (Fig. 2F).

Overall, these results indicate the possibility of using light-ion bombardment to highly enhance the ferroelectric polarization in HZO through defect engineering. Furthermore, an increase in the number of defects can occur owing to light-ion bombardment (figs. S6 and S7). Presumably, the enhanced ferroelectricity of the HZO thin film could be ascribed to the increased defect density inside the HZO thin film by the posttreatment of the He ion bombardment. Regardless, because the enhancement of ferroelectric polarization through defect engineering can be associated with several mechanisms, such as phase transition to the ferroelectric phase (*16*), we explored the origin of the ferroelectric polarization enhancement.

Mechanism of highly enhanced ferroelectricity based on defect engineering

To determine the dominant defect formed and the cause of the enhanced ferroelectricity in the treated HZO thin film, we performed a chemical analysis on the HZO samples with He ion doses of 5×10^{15} and 10^{17} ions/cm^2, respectively, and compared those with the case of the pristine HZO thin film. The pristine HZO and the two HZO thin films with different He ion doses are denoted as pristine (P-HZO), 5×10^{15} ions/cm^2 (L-HZO), and 10^{17} ions/cm^2 (H-HZO), respectively. We constructed EDX elemental maps of Hf (red), Zr (yellow), O (green), and Si (cyan) for the three HZO samples of P-HZO, L-HZO, and H-HZO, respectively (Fig. 3A). The annular dark field (ADF) STEM images and the corresponding EDX maps confirm that the thicknesses of the HZO thin films are ~10 nm but appear to become larger with an increasing dose of He ions, which indicates an increase in molar volume. Comparing the high-resolution high-angle (HA)ADF STEM images of the three HZO samples (Fig. 3B), we observed layer thickening and structural disorder (seen as the increased fuzzy contrast in the image) in the H-HZO sample. This characteristic volume change is also apparent at the atomic scale in the comparison of planar spacings of the orthorhombic phase observed in the three HZO samples (fig. S8). The EDX maps of the three samples revealed that Hf and Zr are sufficiently mixed without elemental segregation, and the distribution of oxygen is wider than

that of Hf and Zr, extending into the Si substrate. This result indicates that a portion of oxygen in the HZO thin film was consumed to form a thin SiO_2 bottom layer on the Si substrate due to the interface reaction during the growth. Thus, the bottom part of the HZO layer would be oxygen deficient as compared to the upper part.

To clarify the structural transition of the HZO thin films to the ferroelectric orthorhombic phase by He ion bombardment, we obtained atomic-scale HAADF STEM images for the three HZO samples and compared them with the simulated counterparts (Fig. 3C). Because the signal contrast in the HAADF imaging mode is approximately scaled to the atomic number squared (Z^{-2}), heavy Hf atoms (Z = 72) appear as bright dots, whereas light oxygen atoms (Z = 8) are hardly observed (*43*). Owing to the polycrystalline nature of the grown HZO thin films, we only observed the atomic structures of the thin films for nanograins aligned along an arbitrary crystal zone axis. Hence, we conducted multiple STEM observations for each sample to determine the polymorphic phase distribution over the film of investigation. As a result, the P-HZO and L-HZO thin films present a mixture of monoclinic and orthorhombic phases, maintaining the monoclinic structure as the dominant phase. We identified the two atomic structures as orthorhombic and monoclinic phases observed in both HZO thin films (figs. S9, A and B). By contrast, we identified the nanograins in the H-HZO thin film as the ferroelectric orthorhombic ($Pca2_1$) phase in all the observations (fig. S9C). We collected an extra dataset confirming this behavior (fig. S10). Interphase boundaries between orthorhombic and monoclinic domains were occasionally spotted in P-HZO films (fig. S11A) but were seldom observed in the orthorhombic grains in H-HZO films. Statistical size measurements indicate that the average grain size was slightly increased by ~2 nm after the He ion bombardment–induced transition to orthorhombic phase (fig. S11B). The HAADF STEM images of the P- and L-HZO samples show the [001]-oriented monoclinic structure, whereas that of the H-HZO samples represents the [110]-oriented orthorhombic structure (Fig. 3C). Position-averaged convergent beam electron diffraction (PACBED) analysis (*44*) was correspondingly performed on the regions of STEM observations to confirm those structures. Consistently, the experimental PACBED patterns show a good match with the simulated PACBED patterns that were derived from the structures identified by the STEM observations (Fig. 3D). Considering the multiple structural observations, we conclude that the He ion bombardment into the pristine HZO thin film readily induces a homogeneous distribution of the ferroelectric orthorhombic phase over the film. These results suggest that atomic defects such as oxygen vacancies and the structural disorder that were induced by He ion bombardment could play an important role in the favorable structural transition to the ferroelectric orthorhombic phase, thus resulting in the notable enhancement of ferroelectricity.

To trace the distribution of oxygen vacancies inside the HZO thin films, we conducted an EELS spectrum imaging (SI) analysis of the O K edge for each HZO film (fig. S12). From the EELS SI datasets, we extracted the averaged O K spectra for the three HZO thin films (see insets in Fig. 3E). The O K energy-loss near-edge structure (ELNES) of HZO presents a notable doublet-peak shape denoted as *a* (peak at ~534 eV) and *b* (peak at ~538 eV) with an energy difference ($\Delta E_{b\text{-}a}$) of ~4 eV, which is attributed to the O 2p orbital hybridized with the Hf (or Zr) d_{e_g} and $d_{t_{2g}}$ orbitals, respectively (*45*). This characteristic peak splitting of the O K ELNES commonly appears in the HfO_2 polymorphs of monoclinic, tetragonal, and cubic phases with slight differences in the energy shift and peak broadening (*45*). Characteristically, the peak sharpness and intensity of peak *a* relative to peak *b* varies with the crystallinity and content of the oxygen vacancies in the HfO_2 film. Poor crystallinity causes a broadening of the double peak structure and also a decrease in the relative intensity of peak *a*, which should be considered when HZO thin films are grown at different annealing temperatures (*46*). Because the HZO samples were grown under the same thermal processing conditions, the most probable origin of the change in the O K ELNES could be the formation of the oxygen vacancies and/or the occurrence of structural disorder due to the He ion bombardment. The notable reduction in the a/b ratio from 0.60 to 0.43 in O K edges (Fig. 3E and fig. S12C) implies that the vacancy defects and structural disorder increased as the dose of He ion bombardment was increased to the optimal value of 10^{17} ions/cm^2 (H-HZO). Because monoclinic and orthorhombic HZO structures have sevenfold coordinate metal ions, a similar broad doublet peak structure of the O K edges is observed (fig. S13A). However, cubic and tetragonal HfO_2 structures with sixfold coordinate metal ions would present substantially sharper double-peaked O K ELNES owing to the higher crystal field symmetry (*46*). Previous studies reported that the formation of oxygen vacancies reduces the first peak (*a*) intensity of the O K doublet edge when other factors are held constant (*46*, *47*). Our density functional theory–based O K-edge simulations for the influence of oxygen vacancies represent the same tendency (figs. S13, B and C, and S14) (*31*).

Therefore, we examined the change in the relative intensity of peak *a* in the O K ELNES to detect the oxygen vacancies inside the HZO thin films and mapped the a/b ratios corresponding to the entire film region by nonlinear least squares (NLLS) fitting analysis (Fig. 3E). The resulting a/b ratio maps show that the oxygen vacancy content is gradually increased from P-HZO to H-HZO. Notably, the entire H-HZO thin film is revealed to be populated by oxygen vacancies, whereas the distribution of the oxygen vacancies is localized and nonuniform in the other two HZO samples (P-HZO and L-HZO); these results are similar to those previously obtained for a conventional HfO_2 sample (*48*). The statistical histograms of the measured a/b ratios for the three HZO samples clearly indicate the varying population behavior of oxygen vacancies from the P- and L-HZO to the H-HZO thin films (fig. S12C). The Gaussian curve fits of the a/b distributions for the three HZO samples demonstrate that the H-HZO thin film is highly oxygen deficient compared with the other two HZO samples. The a/b ratio can also be decreased to some extent as the crystallinity becomes poor (*46*). Given the increase in structural disorder by the He ion bombardment, it is also expected that the a/b ratio would be correspondingly further reduced, making quantification of the vacancy content by EELS difficult.

Theoretical calculations predicted that the injection of oxygen vacancies and substitutional doping into the HfO_2 thin film can promote a phase transition from the monoclinic ($P2_1/c$) to the ferroelectric orthorhombic ($Pca2_1$) phase (*49*, *50*). However, this doping was presumably insufficient to stabilize the metastable ferroelectric phase over the predominant monoclinic phase (*51*, *52*). Experimental work demonstrated that the role of oxygen vacancy is more important in stabilizing the ferroelectric phase, prompting efforts for developing a facile route to controlling the concentration of the oxygen vacancy in the hafnia film (*23*, *30*). On the basis of the STEM-EELS results, we confirmed that He ion bombardment into the HZO thin film efficiently triggers the formation of oxygen vacancy and increases its density and distribution with an increasing He ion dose before causing structural damage (≥ ~10^{18} ions/cm^2). Thus, to explore why the ferroelectric orthorhombic phase is stabilized as the oxygen vacancies are gradually introduced, we surveyed the change in total energy difference between the monoclinic and orthorhombic phases up to the vacancy content of 25% (fig. S13D) (*31*). Unexpectedly, we found that a threshold concentration (>20%) of oxygen vacancy is required to fully stabilize the ferroelectric orthorhombic phase. Annular bright-field STEM imaging confirmed that the oxygen-deficient ferroelectric orthorhombic structure is stably sustained (fig. S13E). A slight reduction in both R-PFM amplitude and remnant piezoresponse at He ion doses below

10^{16} ions/cm^2 may also be associated with the threshold concentration.

Nevertheless, because ferroelectric behavior can have multiple causes, we further analyze the possible thermodynamic mechanisms, including a direct change in molar volume, emergence of ferroelectricity, vacancy formation, and vacancy redistribution. The free energy contribution G_{PE} describing the ferroelectric order-disorder–type transition can be expressed using the macroscopic polarization vector components (*28*):

$$G_{PE} = N_d \left\{ -\frac{J}{2}p^2 - \frac{J_{\mathrm{nl}}}{4}p^4 + \frac{k_{\mathrm{B}}T}{2}\sum_{i=1}^{3} [(1+p_i)\ln(1+p_i) + (1-p_i)\ln(1-p_i)] - d \cdot p_i\left(E_i^e + \frac{E_i^d}{2}\right)\right\} + \frac{\delta_{ijkl}}{2}P_d^2\frac{\partial p_i}{\partial x_k}\frac{\partial p_j}{\partial x_l} \quad (1)$$

where PE denotes electrical polarization; the polarization components are normalized as $p_i = \frac{P_i}{P_d}$, $P_d \cong N_d d$; d is the concentration-independent average value of an elementary dipole moment; the concentration N_d can be related with the average concentration $\langle \delta N_d(r) \rangle$ of different vacancies and/or other defects; J is the effective interaction constant that is negative for the considered case of improper ferroelectricity; J_{nl} is the effective nonlinearity coefficient; δ is the positive gradient coefficient determined by the correlation between the dipole moments and by the contribution of the spatial gradients induced by the surface, as well as by the inhomogeneities inherent to the film; and k_{B} is the Boltzmann constant. We consider that the film surfaces are defect blocking, so all O, Hf, and Zr vacancies inside the film are time independent. Consequently, the average concentration of the vacancies is also time independent and equal to the average value calculated in the initial moment of time. Inserting elastic fields, created by different defects, into the (Hf,Zr)O_2 Gibbs potential and its subsequent expansion on the polarization powers leads to renormalization of the coefficient α by the electrostriction coupling with the Vegard expansion tensor, and the depolarization field effect [see (*28*, *31*)]. Then, a local polar state can emerge for an out-of-plane polarization component $p \equiv \frac{P_3}{P_d}$ when the average concentration of defects $\langle \bar{N}_{\mathrm{d}} \rangle$ is high enough, and the positive terms proportional to $\Lambda/h\varepsilon_0$ and $\delta/\lambda h$ are small enough, where Λ is the effective screening length produced by the exposed free surface and/or imperfect bottom electrode; h is the (Hf,Zr)O_2 film thickness; and ε_0 is the universal dielectric constant.

We show typical dependences of the spontaneous out-of-plane polarization p on the film thickness and vacancy concentration N_{do} (Fig. 4, A and B). The p value decreases rapidly with increasing film thickness h (Fig. 4A). Further, the p value increases rapidly and then saturates with increasing vacancy concentration N_{do} (Fig. 4B). The appearance of ferroelectricity is a typical size-induced effect because the second (positive) and third (negative) terms in eq. S10 for $k = 3$ are thickness dependent. In particular, the third term increases with h as $\frac{1}{h}$, because $\langle \bar{N}_d \rangle = N_{d0}\frac{h_d}{h}(1 - \exp(-\frac{h}{h_d}))$ in accordance with eq. S9, where h_{d} is the decay factor of the defect concentration. Small second and large third terms are required for the negative $\langle\langle \alpha_{33}^R \rangle\rangle$. Making $\Lambda < 0.1$ nm corresponds to a high screening degree by, e.g., either conducting electrodes or ambient free carriers, and using a small gradient coefficient (i.e., $\delta \leq 10^{-10}$ C^{-2} m^3 J) makes the positive terms appear smaller than the negative one. Thus, the third term explains the size effect (Fig. 4, A and B). Assuming that the polar state occurs at $N_{do} \cong 10^{25}$ to 10^{26} m^{-3}, we need to minimize exposure to He; $N_{\mathrm{He}} \cong 10^{14}$ ions/cm^2 can create such a defect concentration over the already available defect "background." Our analysis suggests that the emergence of ferroelectricity in He ion–bombarded hafnia can be driven by a synergy of several factors. In accordance with eq. S10, emergence of ferroelectricity via renormalization of $\langle\langle \alpha_{33}^R \rangle\rangle$ requires a dominant third term. This in turn requires high negative values of electrostriction coupling and large product $\langle W_{33}^d \rangle \langle \bar{N}_d \rangle$. Of these, the electrostriction coefficient q_{33} is of the same order for most ferroics. Hence, the product of the defect concentration and Vegard expansion coefficient $\langle W_{33}^d \rangle$ emerges as a predominant factor controlling ferroelectricity in this material. Although we consider a quasi-equilibrium static configuration, the modeling charge defect transport in HfO_2-based ferroelectrics in a nonequilibrium state could be an interesting further work for electrical cycling.

Finally, although the STEM and theoretical results showed that polarization enhancement based on defect redistribution and phase transition could be one of the possible mechanisms, electrical measurements are still necessary because nonferroelectric factors such as electrostatic effects might contribute to the obtained PFM signal (Figs. 1 and 2). Thus, we directly observed the remnant polarization charge of the electrode sample through an AFM-based positive-up-negative-down (i.e., AFM-PUND) method (*53*, *54*). The remnant polarization charge density clearly increased after the He ion bombardment at a dose of 10^{17} ions/cm^2 (Fig. 4, C and D).

Conclusions

We have demonstrated the high enhancement of ferroelectricity in HZO thin films based on defect engineering by He ion bombardment. Our PFM results indicate that the He ion bombardment substantially enhanced the ferroelectricity in the HZO thin film. In particular, the resonance PFM results indicate that the amplitude of piezoresponse in the ion-bombarded region of 10^{17} ions/cm^2 increased by approximately twofold compared with that in the pristine region. This implies that the fine modulation of light-ion bombardment can substantially enhance the polarization of HZO. Furthermore, the STEM and EELS results show that the large enhancement of ferroelectricity by ion bombardment can be caused by a homogeneous distribution of oxygen vacancies and a phase transition to the ferroelectric phase. Furthermore, the theoretical calculations for thermodynamic mechanism allow us to explore possible mechanisms for the enhancement of the polarization by ion bombardment. These results, regarding the variation of ferroelectricity through defect engineering based on ion bombardment, suggest additional possibilities for ferroelectricity enhancement in HfO_2-based ferroelectrics, including HZO. Thus, He ion bombardment can be a facile technique for stabilizing the ferroelectric orthorhombic phase. Furthermore, this study suggests that highly enhanced ferroelectricity can be selectively obtained by controlling light-ion bombardment with a high spatial resolution. Ultimately, this approach can be directly applied to a semiconductor process without structural modification and, thus, can increase its applicability in next-generation electronic devices, such as ultrascaled ferroelectrics-based transistors and memories.

REFERENCES AND NOTES

1. M. Coll *et al.*, *Appl. Surf. Sci.* **482**, 1–93 (2019).
2. S. Mathews, R. Ramesh, T. Venkatesan, J. Benedetto, *Science* **276**, 238–240 (1997).
3. M. W. Si *et al.*, *Nat. Electron.* **2**, 580–586 (2019).
4. C. H. Ahn *et al.*, *Science* **276**, 1100–1103 (1997).
5. J. Hoffman *et al.*, *Adv. Mater.* **22**, 2957–2961 (2010).
6. J. F. Scott, *Science* **315**, 954–959 (2007).
7. J. F. Ihlefeld *et al.*, *J. Am. Ceram. Soc.* **99**, 2537–2557 (2016).
8. J. Junquera, P. Ghosez, *Nature* **422**, 506–509 (2003).
9. F. Ambriz-Vargas *et al.*, *ACS Appl. Mater. Interfaces* **9**, 13262–13268 (2017).
10. T. S. Böscke, J. Müller, D. Bräuhaus, U. Schröder, U. Böttger, *Appl. Phys. Lett.* **99**, 102903 (2011).
11. M. H. Park *et al.*, *Adv. Mater.* **27**, 1811–1831 (2015).
12. U. Schroeder *et al.*, *Jpn. J. Appl. Phys.* **58** (SL), SL0801 (2019).
13. C. Wang, H. Qiao, Y. Kim, *J. Appl. Phys.* **129**, 010902 (2021).
14. Y. Wei *et al.*, *Nat. Mater.* **17**, 1095–1100 (2018).
15. S. S. Cheema *et al.*, *Nature* **580**, 478–482 (2020).
16. M. Pešić *et al.*, *Adv. Funct. Mater.* **26**, 4601–4612 (2016).
17. T. Y. Lee *et al.*, *ACS Appl. Mater. Interfaces* **11**, 3142–3149 (2019).
18. A. Chouprik *et al.*, *ACS Appl. Electron. Mater.* **1**, 275–287 (2019).
19. H. J. Kim *et al.*, *Nanoscale* **8**, 1383–1389 (2016).
20. P. Buragohain *et al.*, *ACS Appl. Mater. Interfaces* **11**, 35115–35121 (2019).
21. H. Yang *et al.*, *Inorg. Chem.* **59**, 5993–5999 (2020).
22. L. Xu *et al.*, *J. Appl. Phys.* **122**, 124104 (2017).
23. A. Pal *et al.*, *Appl. Phys. Lett.* **110**, 022903 (2017).
24. M. H. Park *et al.*, *Nanoscale* **10**, 716–725 (2018).
25. J. Hur *et al.*, *J. Mater. Res.* **36**, 1206–1213 (2021).
26. K. Y. Chen, P. H. Chen, R. W. Kao, Y. X. Lin, Y. H. Wu, *IEEE Electron Device Lett.* **39**, 87–90 (2018).

27. S. E. Reyes-Lillo, K. F. Garrity, K. M. Rabe, *Phys. Rev. B Condens. Matter Mater. Phys.* **90**, 140103 (2014).
28. M. D. Glinchuk *et al*., *J. Alloys Compd.* **830**, 153628 (2020).
29. P. Nukala *et al*., *Science* **372**, 630–635 (2021).
30. M. Hoffmann *et al*., *J. Appl. Phys.* **118**, 072006 (2015).
31. See the Supplementary Materials.
32. G. S. Was *et al*., *Nucl. Instrum. Methods Phys. Res. B* **412**, 58–65 (2017).
33. M. J. F. Healy, *Nucl. Instrum. Methods Phys. Res. B* **129**, 130–136 (1997).
34. D. Seol *et al*., *Nano Energy* **79**, 105451 (2021).
35. D. Seol, B. Kim, Y. Kim, *Curr. Appl. Phys.* **17**, 661–674 (2017).
36. N. Balke *et al*., *Nanotechnology* **27**, 425707 (2016).
37. Q. N. Chen, Y. Ou, F. Y. Ma, J. Y. Li, *Appl. Phys. Lett.* **104**, 242907 (2014).
38. D. O. Alikin *et al*., *IEEE Trans. Dielectr. Electr. Insul.* **27**, 1478–1485 (2020).
39. S. Jesse, S. V. Kalinin, *J. Phys. D Appl. Phys.* **44**, 464006 (2011).
40. L. Collins, Y. Liu, O. S. Ovchinnikova, R. Proksch, *ACS Nano* **13**, 8055–8066 (2019).
41. A. Labuda, R. Proksch, *Appl. Phys. Lett.* **106**, 253103 (2015).
42. Y. Kim *et al*., *ACS Nano* **6**, 491–500 (2012).
43. S. J. Pennycook *et al*., *J. Anal. Sci. Technol.* **9**, 11 (2018).
44. J. M. Lebeau, S. D. Findlay, L. J. Allen, S. Stemmer, *Ultramicroscopy* **110**, 118–125 (2010).
45. T. Mizoguchi, M. Saitoh, Y. Ikuhara, *J. Phys. Condens. Matter* **21**, 104212 (2009).
46. G. D. Wilk, D. A. Muller, *Appl. Phys. Lett.* **83**, 3984–3986 (2003).
47. S. Stemmer, Z. Q. Chen, W. J. Zhu, T. P. Ma, *J. Microsc.* **210**, 74–79 (2003).
48. C. Li *et al*., *Nano Res.* **8**, 3571–3579 (2015).
49. Y. Zhou *et al*., *Comput. Mater. Sci.* **167**, 143–150 (2019).
50. H. J. Lee *et al*., *Science* **369**, 1343–1347 (2020).
51. R. Batra, T. D. Huan, J. L. Jones, G. Rossetti Jr., R. Ramprasad, *J. Phys. Chem. C* **121**, 4139–4145 (2017).
52. C. Künneth, R. Materlik, M. Falkowski, A. Kersch, *ACS Appl. Nano Mater.* **1**, 254–264 (2018).
53. O. Kwon *et al*., *Small Methods* **5**, e2100781 (2021).
54. O. Kwon *et al*., *Adv. Mater.* **30**, 1703675 (2018).
55. O. Kwon, D. Seol, H. Qiao, Y. Kim, *Adv. Sci. (Weinh.)* **7**, 1901391 (2020).

ACKNOWLEDGMENTS

Funding: This work was supported by the National Research Foundation of Korea (NRF) funded by the Korean government (MSIT) (grant nos. 2021R1A2C2009642, 2020R1F1A1072355, and 2020R1A2C1006207). This work was also supported by the Samsung Advanced Institute of Technology, Samsung Electronics, in Korea. The work of A.N.M. is supported by the National Academy of Sciences of Ukraine (grant no. 1230) and the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement no. 778070. A portion of the work, including He ion exposure, AFM measurements, and data analysis, was conducted at the Center for Nanophase Materials Sciences, which is a US Department of Energy Office of Science User Facility (CNMS2018-219; CNMS2021-B-00917, CNMS2022-R-01082). J.L acknowledges the support of the Korea Basic Science Institute (National Research Facilities and Equipment Center) through a grant funded by the Ministry of Education (no. 2021R1A6C101A429). S.-Y.C. acknowledges the support of the Global Frontier Hybrid Interface Materials of the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2013M3A6B1078872). The use of the TEM instrument (ARM300F) was supported by Advanced Facility Center for Quantum Technology at SKKU. **Author contributions:** Conceptualization: Y.K. Data Curation: S.K., W.-S.J., A.N.M., Y.K. Formal Analysis: S.K., W.-S.J., A.N.M., C.W., E.A.E. Funding acquisition: S.-Y.C., J.L., S.V.K., Y.-M.K., Y.K. Investigation: S.K., W.-S.J., A.N.M., O.K., Y.J., Y.-H.K., H.B., C.W., S.-H.Y., E.A.E., Y.P., K.-J.G., S.-Y.C., J.H.J., J.L., Y.K. Methodology: J.H., S.V.K., Y.-M.K., Y.K. Project administration: S.V.K., Y.-M.K., Y.K. Resources: A.B., S.R., L.C., S.J., M.-H.J., H.W.C., S.K., H.Y.J., O.S.O. Supervision: J.H., S.V.K., Y.-M.K., Y.K. Validation: S.K., W.-S.J., A.N.M., S.V.K., Y.-M.K., Y.K. Writing – original draft: S.K., W.-S.J., C.W., A.N.M., S.V.K., Y.-M.K., Y.K. Writing – review & editing: S.K., W.-S.J., A.N.M., A.B., L.C., O.S.O., J.H., S.V.K., Y.-M.K., Y.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are available in the main text or the supplementary materials. The theoretical calculations for thermodynamic mechanisms were performed and visualized in Mathematica 12.2 Wolfram Research software. Mathematica notebook is available per reasonable request.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk3195
Materials and Methods
Supplementary Text
Figs. S1 to S15
References (*56*–*86*)

Submitted 6 July 2021; resubmitted 30 December 2021
Accepted 7 April 2022
10.1126/science.abk3195

REPORTS

NANOCHEMISTRY

Ultrafast water permeation through nanochannels with a densely fluorous interior surface

Yoshimitsu Itoh[1,2]*, Shuo Chen[1]†, Ryota, Hirahara[1]†, Takeshi Konda[1]†, Tsubasa Aoki[1], Takumi Ueda[3], Ichio Shimada[3,4], James J. Cannon[5,6], Cheng Shao[5], Junichiro Shiomi[5], Kazuhito V. Tabata[7], Hiroyuki Noji[7], Kohei Sato[1]*‡, Takuzo Aida[1,8]*

Ultrafast water permeation in aquaporins is promoted by their hydrophobic interior surface. Polytetrafluoroethylene has a dense fluorine surface, leading to its strong water repellence. We report a series of fluorous oligoamide nanorings with interior diameters ranging from 0.9 to 1.9 nanometers. These nanorings undergo supramolecular polymerization in phospholipid bilayer membranes to form fluorous nanochannels, the interior walls of which are densely covered with fluorine atoms. The nanochannel with the smallest diameter exhibits a water permeation flux that is two orders of magnitude greater than those of aquaporins and carbon nanotubes. The proposed nanochannel exhibits negligible chloride ion (Cl^-) permeability caused by a powerful electrostatic barrier provided by the electrostatically negative fluorous interior surface. Thus, this nanochannel is expected to show nearly perfect salt reflectance for desalination.

The fluorous polymer polytetrafluoroethylene (PTFE) provides a distinct superhydrophobic surface on which densely packed carbon–fluorine (C–F) bonds repel water (*1*). Nevertheless, individual C–F bonds are polar and can interact electrostatically with polar functional groups to form hydrogen bonds (H-bonds) (*2*). This notable bilateral characteristic—known as polar hydrophobicity (*3*)—can be attributed to the properties of fluorine, which include the largest electronegativity and the second smallest atomic size among all elements. The interaction between fluorous surfaces and water has been widely investigated (*4*). Recent reports based on Raman spectroscopy showed that water clusters in the vicinity of fluorous compounds break to yield many hydroxy dangling bonds, whereas considerably fewer dangling bonds are generated with hydrocarbon analogs (*5*, *6*). This observation indicates that a nanochannel with a PTFE-like superhydrophobic interior surface can suppress the formation of water clusters that likely diffuse more sluggishly than nonclustered water molecules.

We developed a series of fluorous oligoamide nanorings ($^{Fm}NR_n$s) that undergo supramolecular polymerization to yield nanochannels ($^{Fm}NC_n$s) with different interior diameters (Fig. 1A). To design $^{Fm}NR_n$s and $^{Fm}NC_n$s, we exploited the polar and hydrophobic natures of C–F bonds. Because a C–F bond is highly polar but is difficult to polarize atomically (*3*), the C–F bond in $^{Fm}NR_n$s can interact electrostatically with adjacent polar amide groups to form H-bonds (*2*). This aspect renders the macrocyclic backbone rigid and causes the C–F bonds to point inward. When $^{Fm}NR_n$s are able to be supramolecularly polymerized in hydrocarbon media to solvophobically form fluorous nanochannels ($^{Fm}NC_n$s) (Fig. 1B),

[1]Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. [2]Japan Science and Technology Agency (JST), Precursory Research for Embryonic Science and Technology (PRESTO), 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. [3]Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [4]RIKEN Center for Biosystems Dynamics Research, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [5]Department of Mechanical Engineering, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. [6]Department of Mechanical Engineering, Faculty of Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan. [7]Department of Applied Chemistry, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. [8]RIKEN Center for Emergent Matter Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.
***Corresponding author. Email: itoh@macro.t.u-tokyo.ac.jp (Y.I.); ksato@macro.t.u-tokyo.ac.jp (K.S.); aida@macro.t.u-tokyo.ac.jp (T.Ai.)** †These authors contributed equally to this work.
‡Present address: School of Life Science and Technology, Tokyo Institute of Technology 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan.

their interior surfaces are densely covered with fluorine atoms and thus can break water clusters (*5*, *6*).

We computationally investigated the influence of the hydrophobicity of the interior nanochannel surface on the water clusters (*7*–*9*) (Fig. 1, C and E). In a molecular dynamics (MD) simulation based on a Lennard–Jones wall with different scaling factors, we changed the hydrophobicity of the interior surface (*10*) (Fig. 1, D and E). As expected from the behavior of the water clusters near the fluorinated compounds (Fig. 1C), we observed the emergence of a larger number of hydroxy dangling bonds as the interior surface became more hydrophobic (Fig. 1D). Subsequently, we investigated the influence of hydrophobicity on the water flow rate. As shown in Fig. 1E, as the hydrophobicity of the wall increased, the flow rate also increased.

A series of fluorous nanorings, $^{F12}NR_4$, $^{F15}NR_5$, $^{F18}NR_6$, and $^{F12}NR_6$ (Fig. 1A), were synthesized through an amide condensation reaction (*10*). Two starting compounds—3,5-diamino-4-fluorobenzamide derivative (**1**) and 2,3-difluoroterephthalic acid (**2**)—were synthesized according to previously reported methods (*10*). Among the considered amide condensation reagents, only pentafluorophenyl diphenylphosphinate (FDPP) (*11*) allowed for compounds **1** and **2** to be connected under microwave irradiation. To isolate $^{F12}NR_4$, $^{F15}NR_5$, and $^{F18}NR_6$ from the reaction mixtures, we conducted reprecipitation, preparative recycling silica gel column chromatography, and gel permeation chromatography to remove large amounts of linear oligomers. The fluorous nanorings were characterized by NMR spectroscopy (figs. S1 to S12) and high-resolution matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry (figs. S16 to S18). We also succeeded in synthesizing the partially fluorinated version ($^{F12}NR_6$) using 3,5-diaminobenzamide and **2** under conditions analogous to those described above (figs. S13 to S15 and S19). The nonfluorous version could not be obtained no matter which optimized conditions were employed.

The core structures of $^{Fm}NR_n$s were optimized by density functional theory (DFT) calculations (figs. S20, S22, S24, and S26). The CPK representation of $^{F12}NR_4$ exhibited a planar conformation (Fig. 2A), whereas the other nanorings exhibited a skewed conformation (Fig. 2, B to D). Although terephthalamide units are rotatable, all the fluorine atoms in the optimized structures were suggested to point inward to form intramolecular "C–F···H–N" H-bonds. The energy gains associated with the intramolecular H-bonding interaction for $^{F12}NR_4$, $^{F15}NR_5$, $^{F18}NR_6$, and $^{F12}NR_6$ were calculated to be –36.2, –47.1, –44.4, and –47.9 kcal mol^{-1}, respectively (figs. S21, S23, S25, and S27). These results indicate that the conformational aspect of $^{Fm}NR_n$s is determined by an H-bonding circular array composed of intramolecular "C–F···H–N" bonds. In this context, when we heated a deuterated dimethyl sulfoxide (DMSO-d_6) solution of $^{Fm}NR_n$s (2.0 mM) from 25° to 65°C, all the signals pertaining to Ar-F in the ^{19}F NMR spectra exhibited a maximum upfield shift of 0.5 parts per million (ppm) (fig. S32, A to D), which is nearly one order of magnitude smaller than those reported for linear analogs of similar

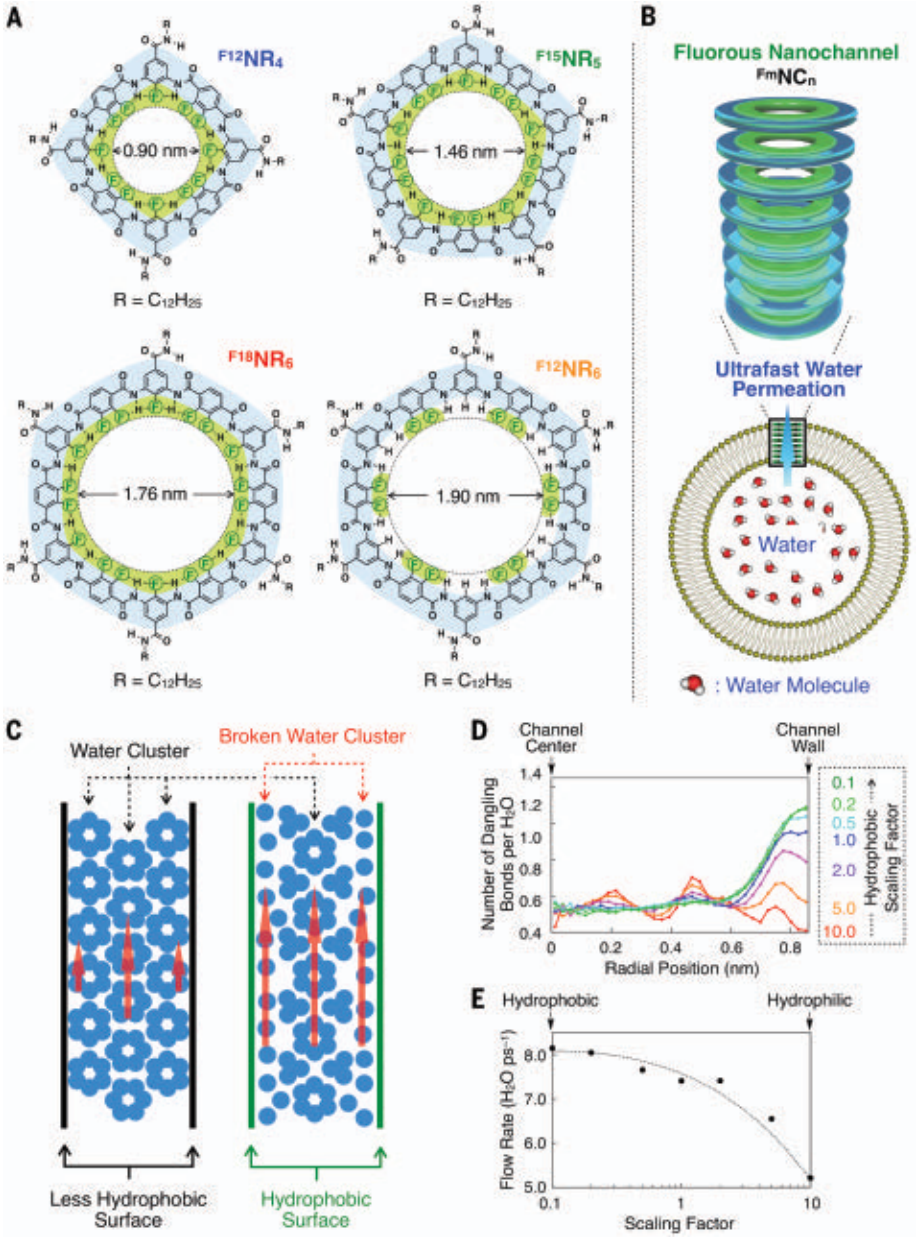


Fig. 1. Series of fluorous nanorings and formation of transmembrane fluorous nanochannels. (**A**) Molecular structures of a series of fluorous nanorings: $^{F12}NR_4$, $^{F15}NR_5$, $^{F18}NR_6$, and $^{F12}NR_6$. The interior diameters of the nanorings were calculated on the basis on the radial water densities obtained by MD simulation of the water permeation through the nanochannel (fig. S37). The wall position was defined as the point at which the water density was 1% of its maximum value. (**B**) Schematic representation of the supramolecularly polymerized $^{Fm}NR_n$ into a fluorous nanochannel ($^{Fm}NC_n$) embedded in a vesicular phospholipid bilayer membrane. (**C**) Schematic representation of the flow of water clusters in hydrophilic and hydrophobic nanochannels. (**D**) Dangling bond distribution of water molecules in a virtual Lennard–Jones (LJ) channel with a diameter of 1.76 nm. The hydrophobicity of the LJ channel was controlled by the scaling factor (*10*). (**E**) Flow rates of water molecules in the LJ channel at different hydrophobicity levels.

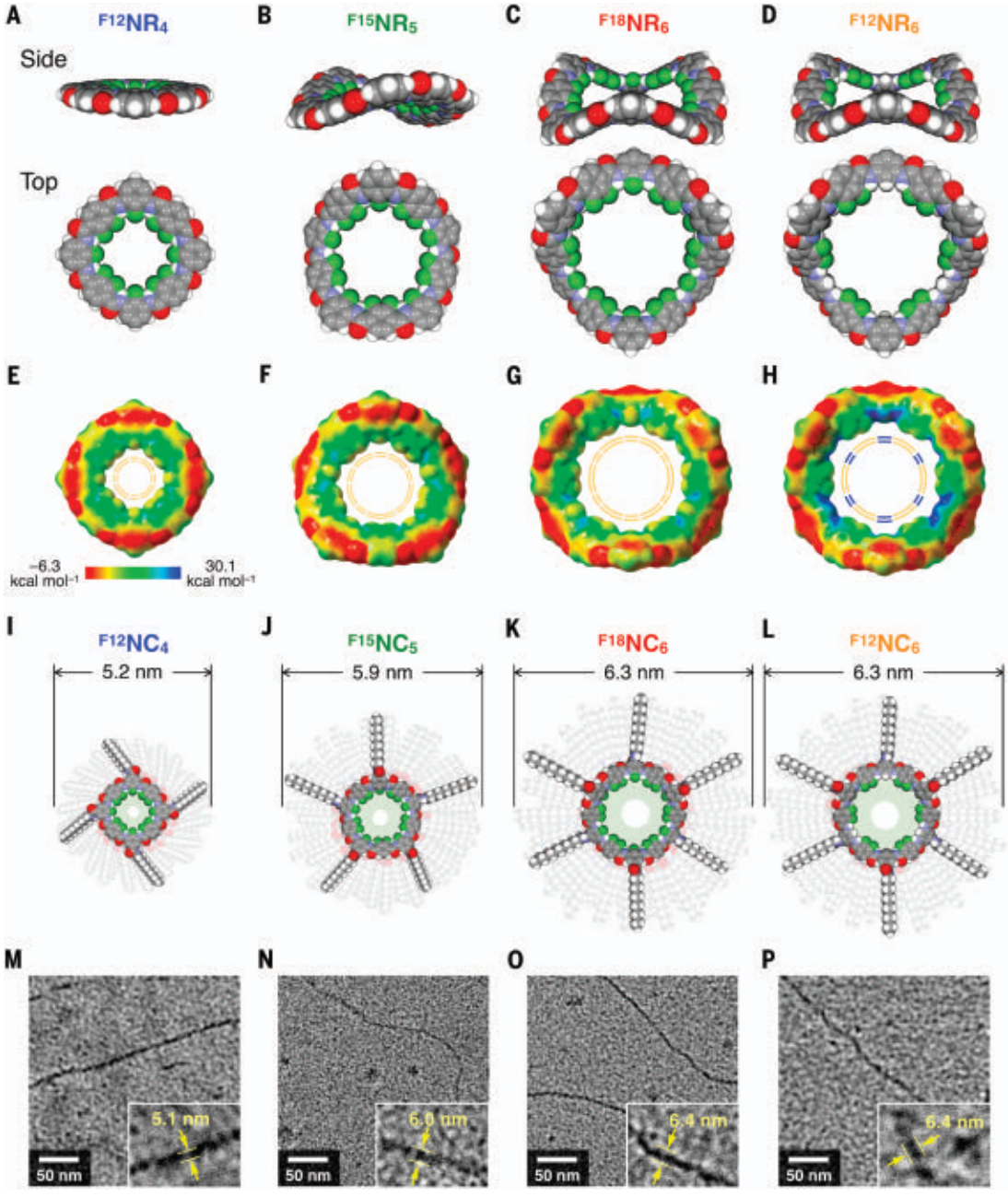


Fig. 2. Formation of fluorous nanochannels. (**A** to **D**) CPK representations of (A) $^{F12}NR_4$, (B) $^{F15}NR_5$, (C) $^{F18}NR_6$, and (D) $^{F12}NR_6$ nanorings obtained by DFT calculations at the B3LYP/6-31G* level of theory (figs. S20, S22, S24, and S26, respectively). (**E** to **H**) Electrostatic potential maps of nanorings (E) $^{F12}NR_4$, (F) $^{F15}NR_5$, (G) $^{F18}NR_6$, and (H) $^{F12}NR_6$ obtained by DFT calculations at the B3LYP/6-31G* level of theory (figs. S20, S22, S24, and S26, respectively). The electrostatic potentials were mapped on the surface with an electron density of 0.0004 units. The red and blue surfaces represent negative and positive regions of the potential, respectively. (**I** to **L**) CPK representations of (I) $^{F12}NC_4$, (J) $^{F15}NC_5$, (K) $^{F18}NC_6$, and (L) $^{F12}NC_6$ nanotubes formed by supramolecular polymerization of $^{F12}NR_4$, $^{F15}NR_5$, $^{F18}NR_6$, and $^{F12}NR_6$, respectively. Each model was constructed on the basis of the structure of $^{Fm}NR_n$ obtained through DFT calculations at the B3LYP/6-31G* level of theory (figs. S20, S22, S24, and S26). The linear alkyl chains were separately optimized and combined with DFT-optimized $^{Fm}NR_n$. (**M** to **P**) TEM micrograph of an air-dried dispersion of supramolecularly polymerized (M) $^{F12}NR_4$, (N) $^{F15}NR_5$, (O) $^{F18}NR_6$, and (P) $^{F12}NR_6$, prepared by heating the suspensions in DMSO/pentadecane at 50°C for 5 min followed by cooling to room temperature. The sample was stained with uranyl acetate.

structures (*2*). Even when the sample solutions were diluted by a factor of 1000 with DMSO-d_6 from 2.0 mM to 2.0 μM at 25°C, no appreciable spectral change was observed (fig. S32, E to H). The observed NMR spectral features were consistent with those obtained computationally. To confirm that the fluorous nanorings $^{F12}NR_4$, $^{F15}NR_5$, $^{F18}NR_6$, and $^{F12}NR_6$ formed nanochannels, their DMSO solutions were added to pentadecane and the mixtures were annealed at 50°C for 5 min and subjected to transmission electron microscopy (TEM) after being allowed to cool to room temperature (*10*). As shown in Fig. 2, M to P, TEM imaging could successfully visualize the presence of extremely long nanotubes with uniform diameters of 5.1, 6.0, 6.4, and 6.4 nm for $^{F12}NR_4$, $^{F15}NR_5$, $^{F18}NR_6$, and $^{F12}NR_6$, respectively, consistent with those estimated from a molecular modeling study (*10*) (Fig. 2, I to L). These results indicated that $^{Fm}NR_n$s (Fig. 1A) can form fluorous nanochannels $^{Fm}NC_n$s in the vesicular phospholipid bilayer membranes (Figs. 1B and 2, I to L) employed for the permeation study.

We performed molecular dynamics simulations of water permeation using the molecular structures of $^{Fm}NC_n$s. The results highlighted that $^{Fm}NC_n$s can permeate water molecules considerably faster than their nonfluorous versions (figs. S35 and S36). To evaluate the water permeability, we initially embedded $^{Fm}NC_n$s in 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) vesicles and used a stopped-flow method in conjunction with light scattering (figs. S38 and S39) to analyze the corresponding water permeation profiles.

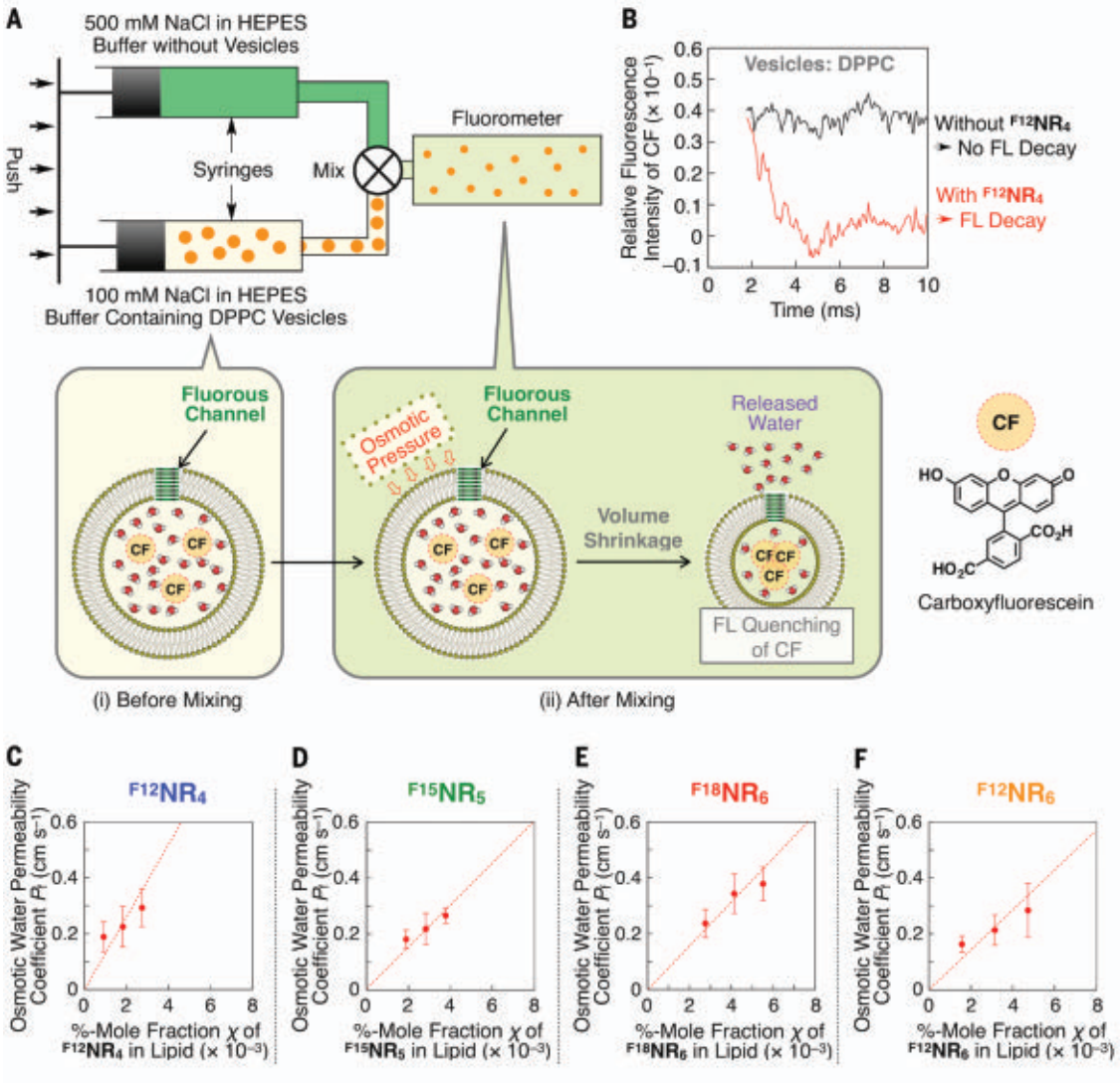


Fig. 3. Stopped-flow fluorescence measurements to evaluate the water permeation kinetics through fluorous nanochannels. (**A**) Schematic representation of the stopped-flow fluorescence measurement setup. A HEPES buffer solution of NaCl ([HEPES] = 10 mM [NaCl] = 500 mM) was added to a HEPES buffer dispersion of CF-enclosing DPPC vesicles ([HEPES] = 10 mM, [NaCl] = 100 mM). Osmotic pressure was generated in the vesicular phospholipid bilayer membrane because the NaCl concentrations in the exterior (300 mM) and interior (100 mM) environments of the vesicle were different. Subsequently, the vesicles were expected to release water molecules through the channels and shrink. This volume shrinkage event was monitored through the fluorescence quenching of CF, owing to its increased concentration in the interior environment of the vesicles. (**B**) Representative stopped-flow fluorescence decay profiles of CF enclosed in DPPC vesicles at 25°C in the presence (red, χ = 0.0018) and absence (black) of $^{F12}NR_4$ in the vesicular phospholipid bilayer membrane. Eight decay profiles were averaged. (**C** to **F**) Dependence of the percent mole fractions (χ) of (C) $^{F12}NR_4$, (D) $^{F15}NR_5$, (E) $^{F18}NR_6$, and (F) $^{F12}NR_6$ in DPPC on the osmotic water permeability coefficients P_f of the corresponding vesicles at 25°C. The error bars represent ± 1 S.D. from the averaged values calculated from 3 to 10 independent experiments.

Although this method has occasionally been used to analyze the water permeation profiles of CNTs (*12*) and AQPs (*13*), we observed a large deviation in the permeation profiles (fig. S38, C to F). Because this deviation was likely caused by spontaneous water leakage through the fluidic vesicular membrane, we used a 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) vesicle instead. Over a wide temperature range—including ambient temperatures—this lipid has been reported to form a less fluidic gel-phase bilayer membrane with a thickness of 3.7 nm (*14*, *15*), which barely permeates water at ambient temperatures (*16*). As observed by the membrane fluidity test based on the laurdan assay (*17*) (fig. S40), the channel-forming $^{Fm}NR_n$s did not deteriorate the structural integrity of the bilayer membrane of DPPC. To enhance the signal-to-noise ratio, we also changed the method of light scattering to fluorescence emission (*18*), which has been used to evaluate the water permeability of AQPs (*19*), nanoring-assembled synthetic nanochannels (*20*), and CNTs (*21*), for example.

As a typical example, a HEPES buffer solution of NaCl ([HEPES] = 10 mM, [NaCl] = 500 mM) was added to a HEPES buffer dispersion of $^{F12}NR_4$-embedded DPPC vesicles ([HEPES] = 10 mM, [NaCl] = 100 mM, [DPPC] = 2.9 mM, [$^{F12}NR_4$] = 0.079 μM). The resulting salt concentrations in the exterior ([NaCl] = 300 mM) and interior ([NaCl] = 100 mM) of the DPPC vesicular membrane were different from each other (Fig. 3A, ii); therefore, osmotic pressure was generated in the membrane, causing the DPPC vesicle to shrink to one-third of its original volume through outward water permeation. We wondered whether osmotic pressure can cause inward NaCl permeation to release pressure. Through a separate set of experiments, we confirmed that the inward NaCl permeation was too slow to affect the evaluation of the outward water permeation (Fig. 4C and fig. S41). With reference to existing methods (*19*–*21*), we employed carboxyfluorescein (CF; Fig. 3A) as an enclosed fluorescence probe for investigating water permeation through fluorous $^{F12}NC_4$ because the fluorescence intensity of CF is concentration-dependent and decreases as a result of self-quenching upon volume shrinkage (fig. S42A). When the phospholipid bilayer membrane at 25°C contained channel-forming $^{F12}NR_4$ at a percent mole fraction χ of 0.0018 (*10*), the

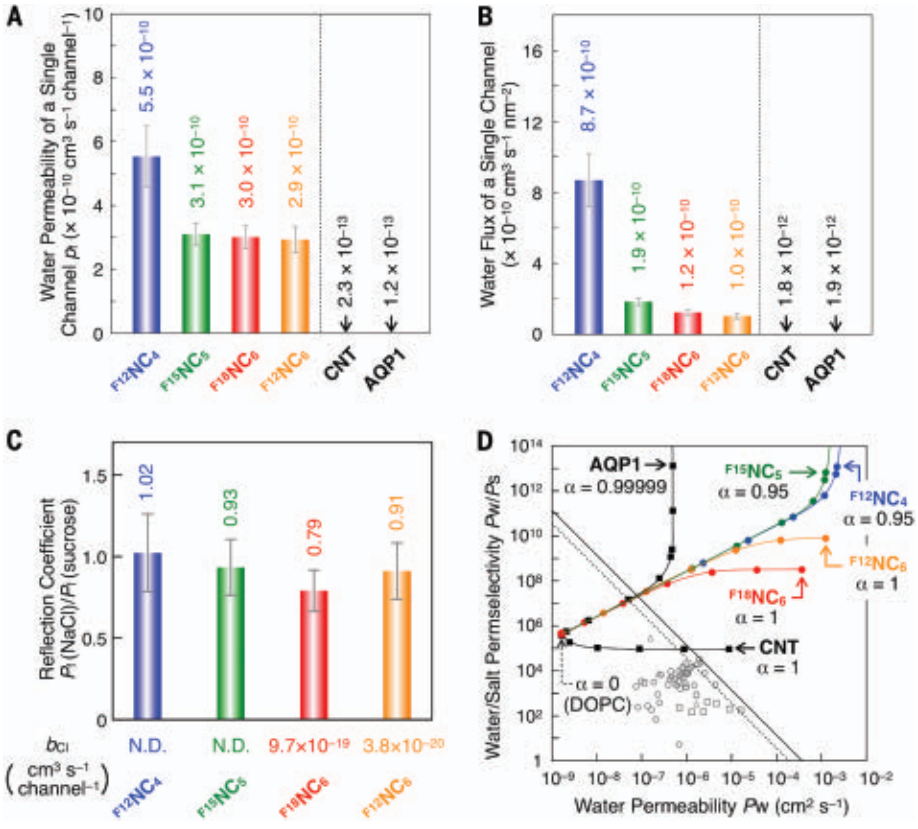


Fig. 4. Water and ion permeabilities through fluorous nanochannels in DPPC vesicles. (**A**) Single-channel osmotic water permeabilities p_f of $^{F12}NC_4$, $^{F15}NC_5$, $^{F18}NC_6$, and $^{F12}NC_6$, along with those of CNT (*21*) and AQP1 (*19*). (**B**) Single-channel fluxes (*f*) of water permeation through $^{F12}NC_4$, $^{F15}NC_5$, $^{F18}NC_6$, and $^{F12}NC_6$, obtained by dividing p_f with their water-permeable cross-sectional areas. The values of CNT (*21*) and AQP1 (*19*) are also shown. (**C**) Reflection coefficients (values in the bar chart) and single-channel Cl^- permeabilities b_{Cl} (below the axis of the bar chart) of $^{F12}NC_4$, $^{F15}NC_5$, $^{F18}NC_6$, and $^{F12}NC_6$. ND represents that the value was under the detection limit. (**D**) Plots of intrinsic water/salt selectivity (P_w/P_s) versus water permeability P_w ($cm^2\ s^{-1}$). P_w/P_s and P_w values for the reconstituted membranes of $^{Fm}NC_n$, CNT, and AQP1 in a DOPC matrix were calculated with a reported method (*27*). The α values in the plots indicate fractional areas of the membranes occupied by the channels. The gray spots are the values for polymeric desalination membranes (*10*). The broken and solid lines in gray represent the trade-off and upper-bound lines, respectively, for the performance of polymeric desalination membranes (*10*).

time-dependent fluorescence intensity change at >500 nm indicated that the CF fluorescence ($[CF]_0$ = 0.5 mM) was quenched within 10 ms (Fig. 3B, red). By contrast, no fluorescence quenching occurred in the absence of channel-forming $^{F12}NR_4$ (Fig. 3B, black), as expected from the lack of water permeability of DPPC vesicles reported previously (*16*). The quenching profile shown in Fig. 3B (red) allowed us to evaluate the kinetic profile of the vesicle shrinkage calibrated by the method shown in fig. S42B and calculate the osmotic water permeability coefficient P_f to be 0.22 cm s^{-1} at χ = 0.0018 (*10*). As shown in Fig. 3C, the P_f values increased linearly as χ increased from 0 to 0.0027, indicating that the vesicle shrinkage rate in the experimental χ range is governed by the amount of water permeated through the channel. Similarly, the P_f values of other fluorous $^{Fm}NC_n$s were obtained as shown in Fig. 3, D to F. To validate that water permeation occurs through the fluorous nanochannel $^{Fm}NC_n$s, we conducted a set of water permeation experiments using $^{F15}NR_5$ in the presence of poly(ethylene glycols) (PEGs) with hydrodynamic diameters (D_h) ranging from 0.4 to 3.2 nm (*22*). As typically observed for interference in water permeation by clogging with polymers, P_f decreased considerably when the D_h of PEG ranged from 1.1 to 1.8, similar to the interior diameter of $^{F15}NR_5$ (1.46 nm) (fig. S43) (*10*, *23*).

The number of $^{F12}NR_4$ molecules per channel (*N*) was suggested to be nearly 10 on the basis of the DFT-optimized stacking distance of $^{F12}NR_4$ (0.41 nm; fig. S28) and membrane thickness of a DPPC vesicle (3.7 nm). The single-channel conductance calculated from this value was consistent with that observed using a planar bilayer membrane under the application of a voltage ranging from 50 to 175 mV without any salt concentration gradient (fig. S45A). Because P_f represents the water permeability per unit area of a phospholipid bilayer membrane, we calculated the single-channel water permeability p_f by using P_f, *N*, χ, and the interior and exterior diameters of the vesicle (Fig. 4A). Unexpectedly, the calculated p_f was 5.5×10^{-10} ($cm^3\ s^{-1}$ per channel), larger than the reported p_f values for the AQP family (AQP1: $p_f = 1.2 \times 10^{-13}$ $cm^3\ s^{-1}$ per channel at 37°C) (*19*), CNT ($p_f = 2.3 \times 10^{-13}$ $cm^3\ s^{-1}$ per channel at room temperature) (*21*), and other synthetic water-permeable nanochannels previously reported (*20*, *24*–*31*) (table S1). Similarly, we evaluated the p_f values of $^{F15}NC_5$, $^{F18}NC_6$, and $^{F12}NC_6$ (Fig. 4A) and noted that these nanochannels exhibited a high water permeability, with the magnitude being the same order as that of $^{F12}NC_4$.

To critically compare the water permeabilities of nanochannels with different channel diameters, we considered the corresponding flux values *f* ($cm^3\ s^{-1}\ nm^{-2}$)—which are normalized permeabilities—taking into account the pore size [i.e., the water-permeable cross-sectional area (*A*), determined using MD simulations (Fig. 1A)] (*10*). The results summarized in Fig. 4B were notable because the thinnest nanochannel, $^{F12}NC_4$ (A = 0.64 nm^2), exhibited the largest *f* value (8.7×10^{-10} $cm^3\ s^{-1}\ nm^{-2}$). The second-thinnest nanochannel, $^{F15}NC_5$, exhibited a slightly larger *f* (A = 1.67 nm^2, 1.9×10^{-10} $cm^3\ s^{-1}\ nm^{-2}$) than those of other nanochannels $^{F18}NC_6$ (A = 2.43 nm^2, 1.2×10^{-10} $cm^3\ s^{-1}\ nm^{-2}$) and $^{F12}NC_6$ (A = 2.83 nm^2, 1.0×10^{-10} $cm^3\ s^{-1}\ nm^{-2}$); however, this value was one-fifth that of $^{F12}NC_4$. The conventional understanding of pipe flow indicates that a thinner channel provides a smaller flux because the shear stress on the interior channel surface is more dominant. However, the *f* values of $^{F12}NC_4$ to $^{F18}NC_6$, as well as those of partially fluorinated $^{F12}NC_6$, indicated the opposite tendency. This anomalous tendency in water flow through nanochannels has been predicted when the water-permeable cross-sectional area of the nanochannel is <0.2 nm^2 (*32*). Computational simulations suggest that water molecules in such confined nanospaces are in alignment and that there is an associated decrease in shear stress on the channel surface. In this context, it is of interest to examine the reasons for this trend even though the cross-sectional areas of our nanochannels are >0.64 nm^2. As shown

in Fig. 1, C to E, fluorous nanochannels with a PTFE-like interior surface can break water clusters (fig. S46) (*5*, *6*). This effect likely becomes more dominant as the diameter of the fluorous nanochannel decreases, leading to the enhancement of water permeation through the nanochannel. The cluster-breaking function is also essential for AQPs and CNTs to permeate water because their interior diameters [~0.3 nm (*33*) and ~0.4 nm (*12*, *21*), respectively] are smaller than those of the water clusters (~0.8 nm) (*34*). Notably, polar amino acid residues and carboxylic acid groups at the entrance parts of AQPs and oxidatively cut CNTs, respectively, have been considered to promote the breakage of water clusters. Notably, the proposed fluorous nanochannels do not need to carry cluster-breaking functional groups at the entrance parts because their interior diameters are larger than those of the water clusters. Instead, the cluster-breaking nature of their interior fluorous surfaces enhances the water permeability, as computationally predicted in the results shown in Fig. 1, D and E, and fig. S46.

A key concern in developing water-permeable nanochannels is salt reflectance, which is important in desalination applications. Although AQPs can realize perfect desalination (*33*), CNTs—the best-performing synthetic nanochannels for fast water permeation—do not effectively reflect salts (*21*). Therefore, we first evaluated the salt reflectance properties of $^{Fm}NC_n$s based on the reflection coefficients (*RC*s) for NaCl. *RC* is the ratio between the water permeability with NaCl ($p_{f,NaCl}$) as an osmolyte to that associated with sucrose ($p_{f,sucrose}$) as a nonpermeable osmolyte (fig. S47), i.e., $p_{f,NaCl}/p_{f,sucrose}$ (*10*). *RC* is unity when the channel does not permeate NaCl (*13*) and smaller than unity when the channel does not reflect NaCl (*28*). The calculated *RC* for the thinnest $^{F12}NC_4$ (A = 0.64 nm^2) was 1.02, indicating a substantially perfect reflection of NaCl (Fig. 4C). However, as the water-permeable cross-sectional area A increased from $^{F12}NC_4$ to $^{F15}NC_5$ (A = 1.67 nm^2) and $^{F18}NC_6$ (A = 2.43 nm^2), *RC* decreased to 0.93 and 0.79, respectively, indicating a negative correlation between A and the ability to reflect NaCl. For partially fluorinated $^{F12}NC_6$ (A = 2.83 nm^2), *RC* was 0.91, similar to that of $^{F15}NC_5$. Overall, all the nanochannels except $^{F18}NC_6$ exhibited excellent salt reflection profiles. We conducted single-channel Cl^- permeability (b_{Cl}, cm^3 s^{-1} per channel) measurements based on a well-established stopped-flow fluorescence technique using channel-embedded DPPC vesicles that enclosed lucigenin, a Cl^--sensitive dye (*21*, *27*, *28*). Cl^- permeability can be approximated to the permeability of salt. In general, charge neutrality is requested in the permeation of salt. Hence, if its cation or anion part does not permeate, the salt is considered to be reflected (*28*). Although the b_{Cl} values of $^{F12}NC_4$ and $^{F15}NC_5$ were smaller than the detection limit, $^{F18}NC_6$ exhibited Cl^--induced fluorescence decay characteristics corresponding to $b_{Cl} = 9.7 \times 10^{-19}$ cm^3 s^{-1} per channel. Although $^{F12}NC_6$ also showed the permeation of Cl^-, the value of b_{Cl} (3.8×10^{-20} cm^3 s^{-1} per channel) was one order of magnitude smaller than that of $^{F18}NC_6$ (Fig. 4C and fig. S48). Meanwhile, we evaluated the Cl^- permeabilities of $^{Fm}NC_n$s using the DOPC vesicles and obtained results that were consistent with those obtained using the DPPC vesicles (fig. S49). Considering the salt reflection (*RC*) and Cl^- permeation experiments, the thinnest $^{F12}NC_4$ exhibited a nearly perfect reflectance, even though its interior diameter (0.9 nm) was larger than the hydration diameter of Cl^- (0.66 nm) (*35*). We calculated the electrostatic potential maps of the four nanorings using the B3LYP/6-31G* level of theory. As shown in Fig. 2, E to G, nanorings $^{F12}NR_4$, $^{F15}NR_5$, and $^{F18}NR_6$ were electrostatically negative at the interior surfaces because of the densely arranged polarized C–F bonds. Thus, the thinnest $^{F12}NC_4$ exhibited a highly powerful electrostatic barrier for the incorporation of Cl^-.

We compared the desalination performances of the proposed fluorous nanochannels with those of reported examples by plotting the water permeabilities (P_w) versus the water/salt selectivities (P_w/P_s) (*10*). As shown in Fig. 4D, the overall desalination performance is higher if the plot of P_w/P_s versus P_w is located more in the top right corner. The P_w/P_s to P_w plots of $^{F12}NC_4$, $^{F15}NC_5$, $^{F18}NC_6$, and $^{F12}NC_6$ are all located in the top right corner of the graph.

We demonstrated that a fluorous nanochannel with a PTFE-like interior surface and an appropriate channel diameter can permeate water at an unprecedentedly high rate with a nearly perfect desalination. These properties originate from the electrostatically negative fluorous interior surface, which can break water clusters to enhance the water permeability and can also provide a powerful electrostatic barrier for the incorporation of Cl^-.

REFERENCES AND NOTES

1. G. J. Puts, P. Crouse, B. M. Ameduri, *Chem. Rev.* **119**, 1763–1805 (2019).
2. C. Li *et al*., *Angew. Chem. Int. Ed.* **44**, 5725–5729 (2005).
3. J. C. Biffinger, H. W. Kim, S. G. DiMagno, *ChemBioChem* **5**, 622–627 (2004).
4. V. H. Dalvi, P. J. Rossky, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13603–13607 (2010).
5. J. R. Robalo *et al*., *J. Am. Chem. Soc.* **141**, 15856–15868 (2019).
6. S. Roy, B. Biswas, N. Ghosh, P. C. Singh, J. A. Mondal, *J. Phys. Chem. C* **123**, 27012–27019 (2019).
7. O. I. Vinogradova, *Int. J. Miner. Process.* **56**, 31–60 (1999).
8. G. Hummer, J. C. Rasaiah, J. P. Noworyta, *Nature* **414**, 188–190 (2001).
9. S. Dalla Bernardina *et al*., *J. Am. Chem. Soc.* **138**, 10437–10443 (2016).
10. Materials and Methods are available as supplementary materials.
11. K. Sato, Y. Itoh, T. Aida, *J. Am. Chem. Soc.* **133**, 13767–13769 (2011).
12. R. H. Tunuguntla *et al*., *Science* **357**, 792–796 (2017).
13. Y. Zhao *et al*., *Environ. Sci. Technol.* **47**, 1496–1503 (2013).
14. B. A. Lewis, D. M. Engelman, *J. Mol. Biol.* **166**, 211–217 (1983).
15. W. G. Hill, R. L. Rivers, M. L. Zeidel, *J. Gen. Physiol.* **114**, 405–414 (1999).
16. M. Jansen, A. Blume, *Biophys. J.* **68**, 997–1008 (1995).
17. F. M. Harris, K. B. Best, J. D. Bell, *Biochim. Biophys. Acta* **1565**, 123–128 (2002).
18. P.-Y. Chen, D. Pearce, A. S. Verkman, *Biochemistry* **27**, 5713–5718 (1988).
19. M. L. Zeidel, S. V. Ambudkar, B. L. Smith, P. Agre, *Biochemistry* **31**, 7436–7440 (1992).
20. X. Zhou *et al*., *Nat. Commun.* **3**, 949 (2012).
21. Y. Li, Z. Li, F. Aydin, J. Quan, X. Chen, Y.-C. Yao, C. Zhan, Y. Chen, T. A. Pham, A. Noy, Water-ion permselectivity of narrow-diameter carbon nanotubes. *Sci. Adv.***6**, eaba9966 (2020).
22. S. Kuga, *J. Chromatogr. A* **206**, 449–461 (1981).
23. X.-Q. Gong, B. J. Nicholson, *Cell Commun. Adhes.* **8**, 187–192 (2009).
24. M. S. Kaucher *et al*., *J. Am. Chem. Soc.* **129**, 11698–11699 (2007).
25. W. Si, P. Xin, Z.-T. Li, J.-L. Hou, *Acc. Chem. Res.* **48**, 1612–1619 (2015).
26. Q. Li *et al*., *Small* **15**, e1804678 (2019).
27. W. Song *et al*., *Nat. Nanotechnol.* **15**, 73–79 (2020).
28. A. Roy *et al*., *Nat. Nanotechnol.* **16**, 911–917 (2021).
29. S. Schneider *et al*., *J. Am. Chem. Soc.* **139**, 3721–3727 (2017).
30. I. Kocsis *et al*., *npj Clean Water* **13**, 1–13 (2018).
31. W. Song, M. Kumar, *Curr. Opin. Chem. Eng.* **25**, 9–17 (2019).
32. J. A. Thomas, A. J. H. McGaughey, *Phys. Rev. Lett.* **102**, 184502 (2009).
33. P. Agre *et al*., *J. Physiol.* **542**, 3–16 (2002).
34. R. Ludwig, *Angew. Chem. Int. Ed.* **40**, 1808–1827 (2001).
35. J. N. Israelachvili, in *Intermolecular and Surface Forces* (Academic press, ed. 3, 2011), pp. 71–90.

ACKNOWLEDGMENTS

We thank A. Kidera from Yokohama City University for valuable discussions. We also thank T. Katagiri from Tokyo University of Technology Graduate School for valuable discussions. Part of the computations were performed using Research Center for Computational Science (Okazaki, Japan). **Funding:** This work was financially supported by a JSPS grant-in-aid for Scientific Research (S) (18H05260) on "Innovative Functional Materials based on Multi-Scale Interfacial Molecular Science" for T.A. Y.I. is grateful for a JSPS Grant-in-Aid for Scientific Research (B) (21H01903) and JST, PRESTO grant JPMJPR21Q1. **Author contributions:** R.H., T.K., S.C., T. Ao., and K.S. designed and performed all the experiments. Y.I. and T. Ai. codesigned the experiments. R.H., T.K., S.C., T. Ao., K.S., Y.I., and T. Ai. analyzed the data and wrote the manuscript. T.U. and I.S. helped with the stopped-flow experiments. J.J.C., C.S., and J.S. performed the MD simulations. K.T. and H.N. obtained the ion conductance measurements. **Competing interests**: The authors declare no competing interests. **Data and materials availability**: All data are available in the manuscript or in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abd0966
Materials and Methods
Figs. S1 to S59
Table S1
References (*36*–*72*)
MDAR Reproducibility Checklist

Submitted 1 June 2020; resubmitted 8 December 2021
Accepted 11 March 2022
10.1126/science.abd0966

WETLAND ECOLOGY

High-resolution mapping of losses and gains of Earth's tidal wetlands

Nicholas J. Murray[1]*, Thomas A. Worthington[2], Pete Bunting[3], Stephanie Duce[1], Valerie Hagger[4], Catherine E. Lovelock[4], Richard Lucas[3], Megan I. Saunders[5], Marcus Sheaves[1], Mark Spalding[6], Nathan J. Waltham[1,7], Mitchell B. Lyons[8]

Tidal wetlands are expected to respond dynamically to global environmental change, but the extent to which wetland losses have been offset by gains remains poorly understood. We developed a global analysis of satellite data to simultaneously monitor change in three highly interconnected intertidal ecosystem types—tidal flats, tidal marshes, and mangroves—from 1999 to 2019. Globally, 13,700 square kilometers of tidal wetlands have been lost, but these have been substantially offset by gains of 9700 km^2, leading to a net change of −4000 km^2 over two decades. We found that 27% of these losses and gains were associated with direct human activities such as conversion to agriculture and restoration of lost wetlands. All other changes were attributed to indirect drivers, including the effects of coastal processes and climate change.

Tidal wetlands are of immense importance to humanity, providing benefits such as carbon storage and sequestration, coastal protection, and fisheries enhancement (*1*, *2*). Unfortunately, intensification of anthropogenic pressure and the growing impacts of climate change are affecting tidal wetlands and their component intertidal ecosystems in pervasive ways. Losses of tidal wetlands are widely reported (*3*–*5*), although at local scales intertidal ecosystems are known to have the capacity to respond to environmental change, gaining extent by means of sediment accumulation, inland migration, and redistribution (*6*–*9*). Redistribution and recovery through natural processes are increasingly supplemented by broad-scale ecosystem restoration activities (*10*). Although a number of studies have suggested that intertidal ecosystems are highly resilient to environmental change (*11*, *12*), little is known about the degree to which gains in tidal wetland extent have counterbalanced known losses. Previous analyses have been unable to address this question due to factors such as focus on mapping single ecosystem types (*9*, *13*, *14*) (which cannot distinguish losses from transitions among adjacent intertidal ecosystems), a lack of consistent data on the global extent and change of tidal marshes (*15*), and uncertainty about the prevailing drivers of tidal wetland change. This has led to considerable uncertainty about how tidal wetlands have changed in recent decades and how they are expected to persist in the future (*11*, *12*).

We report an integrated, globally consistent analysis of the distribution and change of Earth's three intertidal ecosystems: tidal flats, tidal marshes, and mangroves (hereafter referred to collectively as "tidal wetlands"; fig. S1). Where they co-occur, these three ecosystems are highly interconnected, with feedback mechanisms among biological and physical components that interact extensively across the systems (*8*). We investigate the spatiotemporal distribution of tidal wetlands globally by applying stacked machine learning classifiers to remotely sensed data to model their occurrence, detect the type and timing of loss and gain events, and assess the drivers of change over the study period (1999 to 2019). The validated dataset was produced by combining observations from 1,166,385 satellite images acquired by the Landsat 5 to 8 missions with environmental data of variables known to influence the distributions of each ecosystem type, including temperature, slope, and elevation (tables S1 and S2). Tidal wetland loss was defined as the replacement of any of the three focal ecosystems with nonintertidal ecosystems at the 30-m pixel scale, with tidal wetland gain defined as their establishment in pixels where they did not occur in 1999. A weighted random sample of detected changes was used to estimate the contribution of direct human impacts versus indirect drivers, such as sea level rise and natural coastal processes, on tidal wetland losses and gains globally (*16*).

Our global dataset reveals that the total observed area of tidal wetlands in 2019 was at least 354,600 km^2 [95% confidence interval (CI): 244,800 to 363,900 km^2]. Tidal wetlands are unevenly distributed across the world's coastlines, with the largest remaining contiguous tracts occurring as deltaic mangroves fringed by extensive tidal flats in the Amazon Delta, the Northern Bay of Bengal, New Guinea, and the Niger Delta (Fig. 1A). Previous estimates of global tidal marsh extent rely on spatial data compilations with large gaps in coverage that lead to underestimates of extent (*15*), limiting their use for estimating global blue carbon stocks (*17*). Our data therefore allow a first empirical estimate of global tidal marsh extent of 90,800 km^2, obtained by subtracting previously derived extent estimates of mangroves (135,900 km^2) and tidal flats (127,900 km^2) from our global tidal wetland area estimate (*9*, *18*). Our estimate of tidal marsh extent represents 25.6% of the total tidal wetland extent mapped in this study and is 65.1% greater than a previously reported minimum global estimate of 55,000 km^2 (*15*). Because our methods are limited in regions higher than 60°N latitude, where tidal marshes and tidal flats are known to occur, this upward revision of global tidal marsh extent should be considered conservative.

We analyzed tidal wetland change over the 20-year study period and found that losses of 13,700 km^2 (95% CI: −16,800 to −8200 km^2) have been substantially offset by the establishment of 9700 km^2 (95% CI: +4900 to +15,700 km^2) of new tidal wetlands that were not present in 1999 (Table 1). Despite wide geographic variation in the occurrence of tidal wetlands globally, many regions showed a consistent pattern of losses being substantially offset by nearby gains (Fig. 1, B and C). This pattern was most pronounced in the world's major river deltas (*19*), where about one-fifth (19.1%) of the area of tidal wetland changes occurred, despite containing only 7.5% of the world's tidal wetland extent (Fig. 2A). We recorded the greatest tidal wetland change in the Ganges-Brahmaputra (1070 km^2) and Amazon deltas (730 km^2), both of which have exhibited increased tidal wetland extent since 1999 (ratio of loss to gain 0.92 and 0.98, respectively). Many deltas have experienced a net increase in total extent over the past three decades as a result of increases in fluvial sediment supply caused by catchment deforestation and increased upland soil erosion (*19*). Our data, however, suggest a net loss of tidal wetlands on deltas globally, though gains of 2100 km^2 alongside losses of −2300 km^2 indicate the considerable dynamism of these systems. The latter have been associated with multiple direct drivers of change, such as conversion to agriculture and aquaculture (*9*, *14*), urban expansion (*20*), and geomorphic changes due to dikes and channel diversions (*20*), together with many indirect drivers, including shoreline erosion (*9*, *20*), compaction, subsidence and sea level rise (*21*), storm-driven

[1]College of Science and Engineering, James Cook University, Townsville, Australia. [2]Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, UK. [3]Department of Geography and Earth Sciences, Aberystwyth University, Aberystwyth, Wales, UK. [4]School of Biological Sciences, The University of Queensland, Brisbane, Australia. [5]Coasts and Ocean Research Program, CSIRO Oceans and Atmosphere, St. Lucia, Australia. [6]The Nature Conservancy, Department of Physical, Earth, and Environmental Sciences, University of Siena, Siena, Italy. [7]TropWATER, Centre for Tropical Water and Aquatic Ecosystem Research, James Cook University, Townsville, Australia. [8]Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia.
*Corresponding author. Email: nicholas.murray@jcu.edu.au

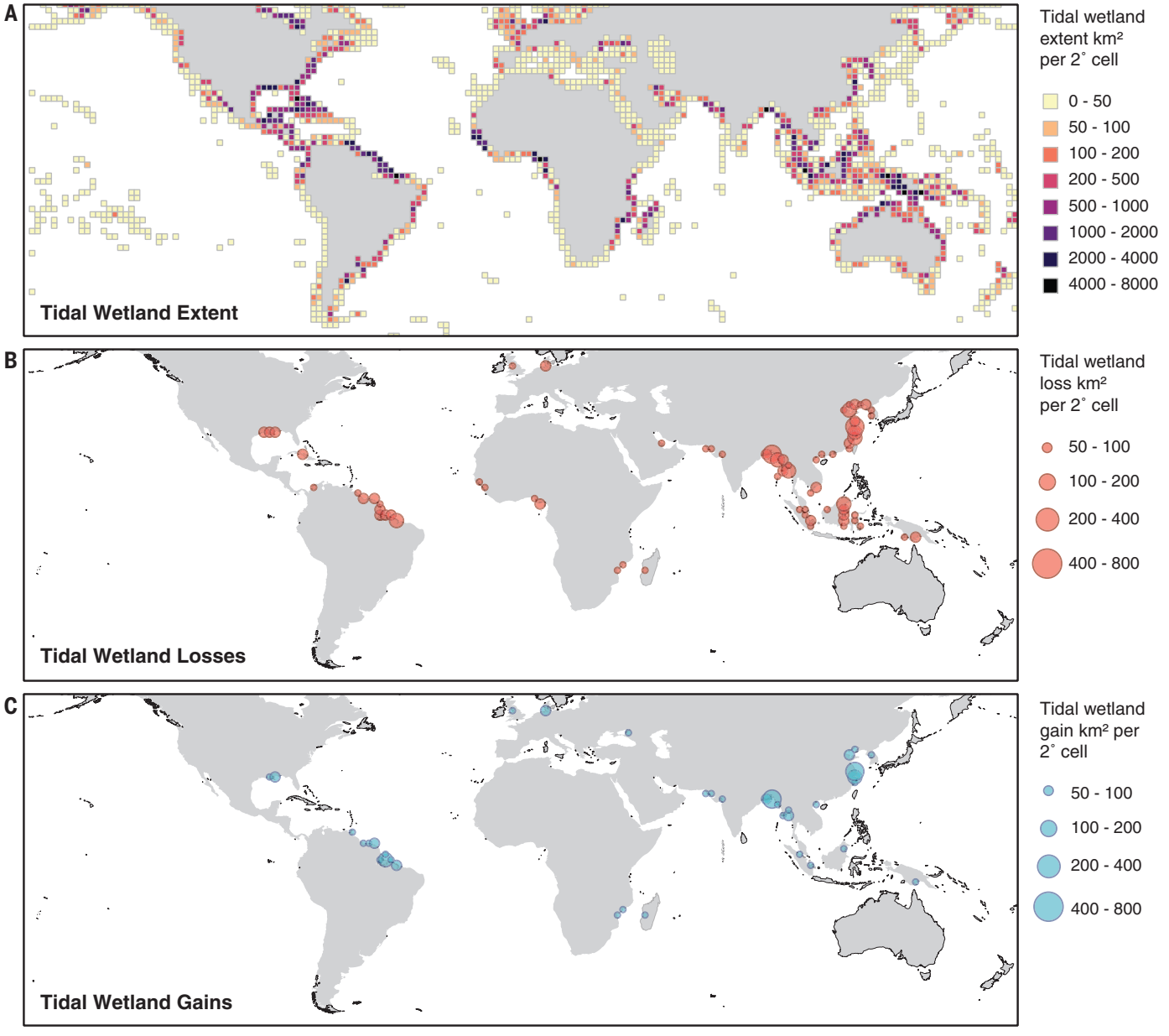


Fig. 1. The global distribution of major tidal wetlands and their changes from 1999 to 2019. (**A**) The 2019 distribution of tidal wetlands is modeled as the combined distribution of the world's three main intertidal ecosystems: tidal flats, tidal marshes, and mangroves. Darker colors indicate greater area of tidal wetlands per 2° grid cell. (**B**) Losses and (**C**) gains over the period 1999 to 2019. Circle sizes indicate the extent of tidal wetland loss and gain over the study period in square kilometers per 2° grid cell.

vegetation loss (*22*), pollution (*23*), and altered sediment supply (*19*).

Of the three intertidal ecosystems included in our analysis, tidal flats experienced both the greatest loss (7000 km^2; 95% CI: 4200 to 8600 km^2) and gain (6700 km^2; 95% CI: 3400 to 10,800 km^2), accounting for almost two-thirds (58.8%) of the total tidal wetland area change (Table 1 and Fig. 3A). A ratio of loss to gain of 1.1 indicates that newly established tidal flats have made a substantial contribution to offsetting the magnitude of their net loss globally. By contrast, mangroves had the highest ratio of loss to gain (3.0), with an estimated net decrease in extent of 3700 km^2 (95% CI: −5400 to −2100 km^2), indicating that extensive mangrove losses have only been partially offset by the 1800-km^2 (95% CI: +900 to +3000 km^2) of new mangroves detected by our analysis. Tidal marshes had the lowest total area change and were the only ecosystem to have a loss to gain ratio of <1, indicating that their gain marginally exceeded their loss for an estimated net increase of 100 km^2 (95% CI: 0 to +100 km^2). Our estimates of global tidal wetland change partitioned by ecosystem type agree in magnitude with recently published estimates of changes in mangrove extent from 2000 to 2016 of −3400 km^2 (*14*). No comparable global estimates of tidal flats and tidal marsh change are available because of differing coverage of change analyses of tidal flats (*9*) and a lack of any data sufficient to support global change analyses of tidal marshes (*15*).

About three-quarters of the 4000 km^2 net global tidal wetland decrease between 1999 and 2019 occurred in Asia (74.1%), with 68.6% concentrated in just three countries: Indonesia

(36%), China (20.6%), and Myanmar (12%). Losses of tropical mangrove forests across Southeastern Asia—particularly Indonesia and Myanmar—are well-documented (*14*, *18*), and the extensive impact of coastal land conversion was confirmed by our analysis (Fig. 3B). In China, there was also a large net reduction of tidal flat extent of >1000 km^2 largely due to reclamation (*24*) but net gains of tidal marshes (+ 200 km^2) that coincided with the rapid expansion of *Spartina alterniflora* across the country's intertidal zone (Fig. 3C) (*25*), thus reducing China's net tidal wetland loss to 800 km^2 (table S3). Outside of Asia, tidal wetlands in Africa had the highest ratio of loss to gain (1.6), indicating a strong loss dynamic that has been associated with severe mangrove degradation, which is most intense in Nigeria, Mozambique, and Guinea-Bissau (Fig. 1B and table S3).

Interpretation of a globally distributed random sample of tidal wetland losses and gains suggested that 39% of losses and 14% of gains were caused by direct human activities (table S9). Direct human activities were defined as observable activities occurring at the location of the detected change (*26*), including conversion to aquaculture, agriculture, plantations, coastal developments, and construction of physical structures such as seawalls and

Table 1. Tidal wetland change estimates by intertidal ecosystem type for different regions of the world from 1999 to 2019. Change estimates are in square kilometers. Tidal wetlands in this study collectively refer to tidal flat, tidal marsh, and mangrove ecosystems, such that area change of tidal wetlands is the sum of the change area of the three component intertidal ecosystems. Per-pixel losses and gains were summarized at these regional scales, with 95% confidence intervals derived from quantitative accuracy assessment in parentheses. Analysis units are realms from the Marine Ecoregions of the World.

Marine ecoregion realm	Tidal flat		Mangrove		Tidal marsh		Tidal wetlands	
	Loss	Gain	Loss	Gain	Loss	Gain	Loss	Gain
Central Indo-Pacific	−952 (−1169 to −570)	1156 (589 to 1871)	−2719 (−3338 to −1626)	758 (387 to 1227)	−12 (−15 to −7)	7 (4 to 12)	−3683 (−4522 to −2203)	1921 (980 to 3110)
Western Indo-Pacific	−1573 (−1932 to −941)	1741 (888 to 2818)	−1139 (−1398 to −681)	380 (194 to 616)	−18 (−22 to −11)	66 (34 to 107)	−2730 (−3351 to −1633)	2187 (1115 to 3541)
Tropical Atlantic	−1067 (−1309 to −638)	1082 (552 to 1752)	−1541 (−1892 to −922)	567 (289 to 919)	−14 (−18 to −9)	5 (2 to 8)	−2622 (−3219 to −1569)	1654 (844 to 2678)
Temperate Northern Pacific	−2355 (−2891 to −1408)	1415 (722 to 2291)	−31 (−38 to −18)	26 (13 to 43)	−192 (−236 to −115)	428 (218 to 693)	−2578 (−3165 to −1542)	1869 (954 to 3027)
Temperate Northern Atlantic	−594 (−729 to −355)	827 (422 to 1338)	−2 (−2 to −1)	2 (1 to 3)	−680 (−835 to −407)	478 (244 to 774)	−1276 (−1566 to −763)	1307 (667 to 2116)
Tropical Eastern Pacific	−106 (−131 to −64)	110 (56 to 178)	−93 (−115 to −56)	86 (44 to 139)	0 (0 to 0)	0 (0 to 1)	−200 (−246 to −120)	196 (100 to 317)
Temperate South America	−110 (−135 to −66)	117 (60 to 190)	−4 (−5 to −2)	5 (3 to 8)	−66 (−82 to −40)	111 (57 to 180)	−181 (−222 to −108)	233 (119 to 378)
Arctic	−118 (−145 to −71)	173 (88 to 280)	0 (0 to 0)	0 (0 to 0)	−30 (−36 to −18)	36 (19 to 59)	−148 (−182 to −88)	209 (107 to 339)
Temperate Australasia	−86 (−105 to −51)	66 (34 to 107)	−13 (−16 to −8)	2 (1 to 2)	−43 (−53 to −26)	20 (10 to 32)	−141 (−174 to −85)	87 (44 to 141)
Eastern Indo-Pacific	−53 (−65 to −32)	4 (2 to 7)	−1 (−1 to −1)	1 (0 to 1)	0 (0 to 0)	0 (0 to 0)	−54 (−67 to −33)	5 (3 to 8)
Temperate Southern Africa	−9 (−11 to −5)	8 (4 to 13)	−19 (−23 to −11)	1 (0 to 1)	−6 (−7 to −4)	11 (6 to 18)	−33 (−41 to −20)	20 (10 to 32)
Southern Ocean	−4 (−5 to −3)	2 (1 to 4)	0 (0 to 0)	0 (0 to 0)	−1 (−1 to −1)	1 (1 to 2)	−5 (−6 to −3)	4 (2 to 6)
Total	−7028 (−8628 to −4204)	6700 (3418 to 10849)	−5561 (−3326 to −6827)	1828 (932 to 2960)	−1064 (−1306 to − 636)	1164 (594 to 1884)	−13652 (−16760 to −8166)	9692 (4944 to 15693)

dikes (*9*) (fig. S8). They also include drivers of gain such as mangrove planting, restoration activities, or coastal modifications to promote tidal exchange (Fig. 2B and fig. S8).

At the continental scale, Asia was identified as the global center of tidal wetland loss from direct human activities (fig. S9). In Asia, direct drivers accounted for more than two-thirds of the losses of each ecosystem (mangrove, 75%; tidal marsh, 69%; tidal flats; 62%; table S10), confirming the negative effects of widespread coastal transformation on coastal ecosystems. Although the impact of coastal development on mangroves and tidal wetlands have been previously reported (*9*, *14*), our results reveal that Asian tidal marshes have similarly been severely degraded by human activities. Compared with Asia, direct human activities had a much lesser role in the losses of tidal wetlands in Europe (28%), Africa (27%), North America (9%), South America (2%), and Oceania (0%; fig. S9).

Indirect or ex situ drivers include both natural coastal processes and those influenced by human activities remotely from the location of observed change. They include processes of isostatic change (*21*), sea level rise (*8*), storm impacts (*22*), erosion and progradation (*22*), along-shore coastal development (*9*), and the combined effects of all of the above. More than 90% of tidal wetland losses in North America (91%), South America (98%), and Oceania (100%) were attributed to indirect drivers (fig. S9). Globally, indirect drivers accounted for most losses of tidal marshes (78%) and tidal flats (66%), whereas mangrove losses were equally a result of direct and indirect drivers (50%; Table S9).

Most tidal wetland gains (86%) were the result of indirect drivers, highlighting the prominent role that broad-scale coastal processes have in controlling tidal wetland extent and facilitating natural regeneration. However, disentangling the specific processes underpinning tidal wetland change is challenging with analyses conducted at large spatial scales. In most cases direct drivers could be clearly identified, but many indirect drivers operate over large spatial and temporal scales and may originate tens to thousands of kilometers from an observed tidal wetland change. Change in ecosystem extent can also be the result of more than one indirect driver or of interactions between drivers. Our work therefore suggests a need for continued monitoring, experiments, and models that can account for these complexities to help characterize and predict global tidal wetland dynamics.

There is potential to use our analysis, which is designed to be periodically updated, to track larger-scale coastal ecosystem restoration activities. Although there has been a surge in coastal restoration efforts worldwide (*27*), many of these have been unsuccessful (*10*). Monitoring progress of restoration remotely, independently, and at broad scales could contribute to reporting on international conservation initiatives such as the UN Decade on Ecosystem Restoration, on targets associated with the Convention on Biological Diversity, and on mitigation commitments made under the UN Framework Convention on Climate Change (*27*). The driver analysis indicated that 14% of observed tidal wetland gains were attributable to direct human interventions (table S9). These activities were most apparent for tidal marshes and were typically the product of site scale restoration activities (Fig. 2B and fig. S8).

Our analysis enables the detection and characterization of dynamic transitions among intertidal ecosystem types globally. Transitions have been linked to a number of physical and climatic factors such as sea level rise, geomorphic changes, and variation in temperature and rainfall (*28*, *29*). We found that 1.9% of the world's tidal wetlands exhibited transitions among ecosystem types over the study

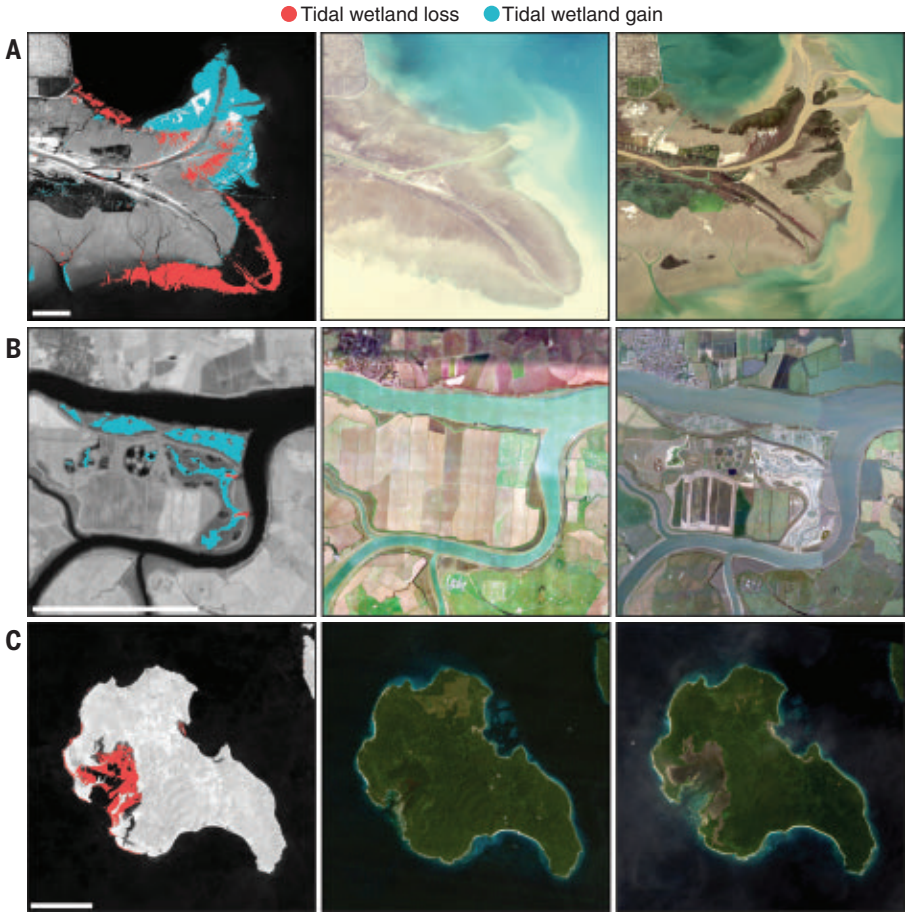


Fig. 2. Representative examples of 1999 to 2019 tidal wetland loss and gain. (**A**) Losses and gains of tidal flats and tidal marshes after diversion of the main channel in 1996 in the Yellow River delta in China (left). Reference images acquired in 1998 (middle) and 2020 (right). (**B**) Tidal marsh gain due to Europe's largest coastal wetland restoration project, UK (left). Reference images acquired in 1999 (middle) and 2018 (right). (**C**) Mangrove loss due to tectonic subsidence after the Aceh-Andaman earthquake of 2004, Katchal Island, Nicobar Islands (left). Reference images acquired 1992 (middle) and 2019 (right). Imagery data are from USGS (A) and (C) and Google Earth Pro (B). All scale bars are 5 km.

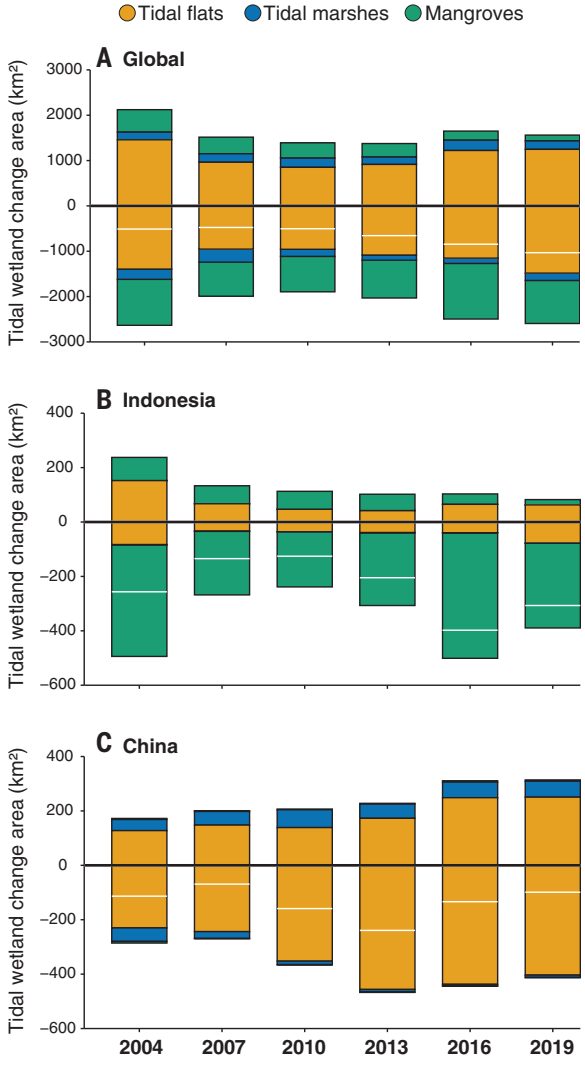


Fig. 3. Tidal wetland change totals from 1999 to 2019. Panels show the losses and gains of tidal wetlands per time step for (**A**) the global study area, (**B**) Indonesia, and (**C**) China. Shaded colors represent intertidal ecosystem types mapped by this analysis. White lines indicate the net change of tidal wetlands. Note that area change cannot be estimated for the initial time step of the analysis (1999 to 2001).

period (6700 km^2; table S4). Transition events tended to be spatially clustered with areas of large losses and gains and in many cases may be linked to the same drivers (*8*). More than 55% of transitions (>3600 km^2) involved colonization of tidal flats by marshes or mangroves and a further 27% consisted of transitions from mangrove to tidal marsh or vice versa (table S4).

Our classifier accurately detected known events of coastal change across the three ecosystem types. For example, the magnitude 9.2 Aceh-Andaman earthquake on 26 December 2004 caused up to 2.9 m of tectonic subsidence in the Andaman and Nicobar Islands, leading to land submergence and a loss of >90% of mangrove extent in some localities (Fig. 2C) (30). However, as for all earth observation-derived estimates of land cover change, there are limitations. These include the spatial resolution of sensor data (which limit the ability of our analysis to detect change in narrow linear features such as waterways), model uncertainty, errors of omission and commission, and a lack of polar coverage. Validation of our data products was effective in characterizing these uncertainties, which were propagated through our estimates of tidal wetland extent and change (tables S5 to S8).

By simultaneously mapping three of Earth's intertidal ecosystems, this work enables a synoptic view of change of three of the world's highly connected intertidal coastal ecosystems. This approach offers an advantage over single-ecosystem mapping studies as short- and long-term dynamic transitions between ecosystem types can cause considerable apparent change in individual ecosystems. Although our study is unable to account for the impact of centuries of anthropogenic coastal transformation and related pressures (*8*), it has established an observational record of recent tidal wetland changes with preliminary attribution of change drivers. Such information has the potential to promote objective monitoring of conservation and restoration efforts, assess the impacts of elevating pressures, guide new studies of changing ecosystem structure, function, and service provision in newly established tidal wetlands (*8*, *29*), and improve our understanding of the resilience of tidal wetlands in the face of global change. In turn, it can support efforts to anticipate the future of global coastal environments and to develop adaptive responses to change.

REFERENCES AND NOTES

1. E. B. Barbier *et al.*, *Ecol. Monogr.* **81**, 169–193 (2011).
2. M. D. Spalding *et al.*, *Conserv. Lett.* **7**, 293–301 (2014).
3. E. B. Barbier, *Science* **345**, 1250–1251 (2014).
4. D. A. Friess *et al.*, *Annu. Rev. Environ. Resour.* **44**, 89–115 (2019).
5. D. M. FitzGerald, Z. Hughes, *Annu. Rev. Earth Planet. Sci.* **47**, 481–517 (2019).
6. C. E. Lovelock *et al.*, *Nature* **526**, 559–563 (2015).
7. S. Fagherazzi *et al.*, *Rev. Geophys.* **50**, RG1002 (2012).
8. M. L. Kirwan, J. P. Megonigal, *Nature* **504**, 53–60 (2013).
9. N. J. Murray *et al.*, *Nature* **565**, 222–225 (2019).
10. E. Bayraktarov *et al.*, *Ecol. Appl.* **26**, 1055–1074 (2016).
11. M. L. Kirwan, S. Temmerman, E. E. Skeehan, G. R. Guntenspergen, S. Fagherazzi, *Nat. Clim. Chang.* **6**, 253–260 (2016).
12. M. Schuerch *et al.*, *Nature* **561**, 231–234 (2018).
13. P. Bunting *et al.*, *Remote Sens.* **10**, 1669 (2018).
14. L. Goldberg, D. Lagomasino, N. Thomas, T. Fatoyinbo, *Glob. Change Biol.* **26**, 5844–5855 (2020).
15. C. J. Mcowen *et al.*, *Biodivers. Data J.* **5**, e11764 (2017).
16. Materials and methods are available as supplementary materials.
17. P. I. Macreadie *et al.*, *Nat. Commun.* **10**, 3998 (2019).
18. M. Spalding, M. Leal, Eds., "The state of the world's mangroves 2021" (Global Mangrove Alliance, 2021).
19. J. H. Nienhuis *et al.*, *Nature* **577**, 514–518 (2020).
20. K. B. Gedan, B. R. Silliman, M. D. Bertness, *Annu. Rev. Mar. Sci.* **1**, 117–141 (2009).
21. J. P. M. Syvitski *et al.*, *Nat. Geosci.* **2**, 681–686 (2009).
22. D. Lagomasino *et al.*, *Nat. Commun.* **12**, 4003 (2021).
23. L. A. Deegan *et al.*, *Nature* **490**, 388–392 (2012).
24. X. Wang *et al.*, *Nat. Sustain.* **4**, 1076–1083 (2021).
25. M. Liu *et al.*, *Remote Sens.* **10**, 1933 (2018).
26. X.-P. Song *et al.*, *Nature* **560**, 639–643 (2018).
27. C. M. Duarte *et al.*, *Nature* **580**, 39–51 (2020).
28. K. C. Cavanaugh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 21602–21608 (2019).
29. N. Saintilan *et al.*, *Science* **368**, 1118–1121 (2020).
30. P. Nehru, P. Balasubramanian, *For. Ecol. Manage.* **427**, 70–77 (2018).
31. N. Murray *et al.*, Code for high-resolution mapping of losses and gains of Earth's tidal wetlands, v1.0.0 Zenodo (2022); https://doi.org/10.5281/zenodo.5968865.
32. N. Murray *et al.*, Intertidal ecosystem training data for mapping Earth's tidal wetlands, Figshare (2022); https://doi.org/10.6084/m9.figshare.19121660.
33. N. Murray *et al.*, Data descriptor for Murray global tidal wetland change, v1.0.0 Zenodo (2022); https://doi.org/10.5281/zenodo.6503080.

ACKNOWLEDGMENTS

We thank R. Canto, M. Toor, and N. Y. Cárdenas for assistance with developing the training data. Landsat data are courtesy of NASA Goddard Space Flight Center and the US Geological Survey. We thank the Google Earth Engine team for providing Earth Engine, and M. Dewitt and N. Clinton for assistance with processing Landsat covariates in Earth Engine. **Funding:** This work was supported by the following: Australian Research Council Discovery Early Career Research Award DE190100101 (N.J.M.); Australian Research Council Discovery Australian Laureate Fellowship FL200100133 (C.E.L.); Julius Career Award from CSIRO (M.I.S.); Australian Research Council Linkage grant LP170101171 (V.H.). **Author contributions:** Conceptualization: N.J.M., T.A.W., and M.B.L. Data curation: N.J.M., P.B., R.L., and T.A.W. Methodology: N.J.M., T.A.W., M.B.L., C.E.L., V.H., and M.I.S. Formal Analysis: N.J.M., T.A.W., M.B.L., and S.D. Funding acquisition: N.J.M. Investigation – N.J.M., T.A.W., P.B., V.H., C.E.L., M.I.S., M.Sh., M.Sp., N.J.W., and M.B.L. Writing – original draft: N.J.M., T.A.W., and M.B.L. with input from all authors. Writing – review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Code to

develop the global tidal wetland change product is available at Zenodo (*31*). Global tidal wetland training data is available at Figshare (*32*). The tidal wetland data products are viewable at www.globalintertidalchange.org and made available in several formats via Zenodo (*33*). Landsat Archive data are freely available courtesy of the U.S. Geological Survey (https://www.usgs.gov/landsat-missions/landsat-collection-1) and via Google Earth Engine Data Catalog (https://developers.google.com/earth-engine/datasets/catalog/landsat).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm9583
Materials and Methods
Figs. S1 to S9
Tables S1 to S10
MDAR Reproducibility Checklist
References (*34*–*75*)

Submitted 26 October 2021; accepted 1 April 2022
10.1126/science.abm9583

ORGANIC CHEMISTRY

Modular access to substituted cyclohexanes with kinetic stereocontrol

Yangyang Li[1], Yuqiang Li[1], Hongjin Shi[1], Hong Wei[1], Haoyang Li[1], Ignacio Funes-Ardoiz[2]*, Guoyin Yin[1]*

Substituted six-membered cyclic hydrocarbons are common constituents of biologically active compounds. Although methods for the synthesis of thermodynamically favored, disubstituted cyclohexanes are well established, a reliable and modular protocol for the synthesis of their stereoisomers is still elusive. Herein, we report a general strategy for the modular synthesis of disubstituted cyclohexanes with excellent kinetic stereocontrol from readily accessible substituted methylenecyclohexanes by the implementation of chain-walking catalysis. Mechanistically, the initial introduction of a sterically demanding boron ester group adjacent to the cyclohexane is key to guiding the stereochemical outcome. The synthetic potential of this methodology has been highlighted in late-stage modification of complex bioactive molecules and in comparison with current cross-coupling techniques.

Three-dimensionally complex geometries frequently confer improved biological activities and physical features to drug molecules compared to their flat bioisosteres (*1*–*3*). Creating efficient synthetic protocols to construct saturated rings in a stereospecific manner has therefore stimulated great interest (*4*–*6*). Cyclohexanes bearing one substituent in the equatorial position and another in the axial position, namely the 1,2-*cis*, 1,3-*trans*, and 1,4-*cis* substitution patterns (Fig. 1A), represent a key scaffold in a wide range of molecules of pharmaceutical interest, such as linrodostat (*7*), an anticancer phase 3 drug; cathepsin S inhibitor (*8*); oxysterols receptor LXR-beta (*9*); and histamine H_3R ADS10227 (*10*) (Fig. 1B). Furthermore, they frequently exhibit better bioactivity than their thermodynamically favored isomers (*10*, *11*). Although studies toward the development of efficient methods for the construction of cyclohexane skeletons have not stopped since the discovery of the Diels-Alder reaction (*12*–*14*), modular and variable approaches for the synthesis of thermodynamically disfavored substituted cyclohexanes are still lacking. Conventional cross-couplings of cyclohexyl derivatives offer a reliable platform to prepare substituted cyclohexanes, but these methods uniformly favor thermodynamically stable stereoisomeric products, that is, the 1,2-*trans*, 1,3-*cis*, and 1,4-*trans* isomers (*15*–*17*). Selective hydrogenation of trisubstituted alkenes (*18*–*22*) or hydrofunctionalization of 1,1-disubstituted alkenes (*23*) would be alternative pathways to access these compounds. However, controlling facial selectivity in these processes is still a challenging task (Fig. 1C, left). Although transition metal–catalyzed hydrogenation of substituted arenes is particularly appealing as a route to all-*cis*-substituted cyclohexanes (*24*–*26*), this strategy does not afford access to cyclohexanes with 1,3-*trans* substituents. Given the prevalence of cyclohexane scaffolds and the limitation of present synthetic methods, development of strategies for the modular synthesis of the thermodynamically disfavored disubstituted cyclohexanes in a highly stereoselective and efficient manner would not only greatly enrich the toolbox of organic synthetic chemists but also enlarge the compound library available for drug discovery.

Inspired by the stereoselectivity induced by the size of the metal hydride reagent in the reduction of substituted cyclohexanones (*27*), as well as our previous studies on migratory difunctionalization of alkenes (*28*–*30*), we speculated that if a bulky group could be incorporated into the exocyclic position of a substituted methylenecyclohexane, it would generate a sterically demanding trisubstituted alkene capable of enhancing stereodiscrimination by a catalyst (Fig. 1C, right). Subsequent coupling with an electrophile would afford the thermodynamically disfavored disubstituted cyclohexane with a good diastereomeric ratio (dr). Herein, we report our successful implementation of this concept by in situ incorporation of a transformable boron group through the use of substituted methylenecyclohexanes as coupling partners in a nickel-catalyzed migratory carboboration reaction (Fig. 1D). The installation of a boron functional group not only plays an important role in controlling the stereoselectivity but also generates a key reactive site for further late-stage functionalization by means of well-known C–B bond functionalization chemistry.

We commenced this study by choosing terminal alkene **1**, bis(pinacolato)diboron (B_2pin_2) (**2**), and organohalide ***3*** as the model substrates. Ligand evaluation demonstrated that the chemoselectivity of this reaction was highly dependent on the ligand framework and that sterically hindered pyridinyl oxazoline (PyrOx) ligands could efficiently promote the formation of 1,4-*cis*-product **4** (Fig. 2A). After carefully examining each reaction parameter, we identified the combination of nickel acetylacetonate [$Ni(acac)_2$] and PyrOx ligand **L6** as precatalyst, lithium methoxide (LiOMe) as base, and *N*-methyl-2-pyrrolidone (NMP)/1,4-dioxane (9:1) as solvent at 35°C as optimal reaction conditions (see supplementary materials for details), which furnished the 1,4-*cis*-product **4** with excellent cis/trans selectivity (>99:1). In contrast, when B_2pin_2 was replaced by pinacolborane (HBpin), anti-Markovnikov hydroalkylation product **8** was obtained in 30% yield, but with a poorer diastereoselectivity (cis/trans = 5:1) (Fig. 2B), while the high level of diastereoselectivity (cis/trans = 83:1) was retained in the hydroalkylation reaction of alkenylboron **5** (Fig. 2C). The diminished reaction efficiency was caused by alkene isomerization, which was probably due to the formation of unexpected Ni(I)-H species during the precatalyst activation (*31*). These results clearly support the hypothesis that the initial incorporation of the pinacol boronic ester (Bpin) group plays an important role in increasing the stereoselectivity of this reaction. To further confirm this effect, density functional theory calculations (see supplementary materials for computational details) were carried out on the Ni-H migratory insertion step (Fig. 2D). After a careful transition-states analysis of the different equatorial and axial conformers of the Ni-H insertion into the alkene **1**, we found that cis migratory insertion is favored by 3.7 kcal/mol owing to the large steric hindrance for the trans migratory insertion between the PyrOx ligand and the boron functional group. Despite the small size of hydrogen, the local steric hindrance in the transition state,

[1]The Institute for Advanced Studies, Wuhan University, Wuhan 430072, China. [2]Department of Chemistry, Centro de Investigación en Síntesis Química, Universidad de La Rioja, 26006 Logroño, Spain.
*Corresponding author. Email: ignacio.funesa@unirioja.es (I.F.-A.); yinguoyin@whu.edu.cn (G.Y.)

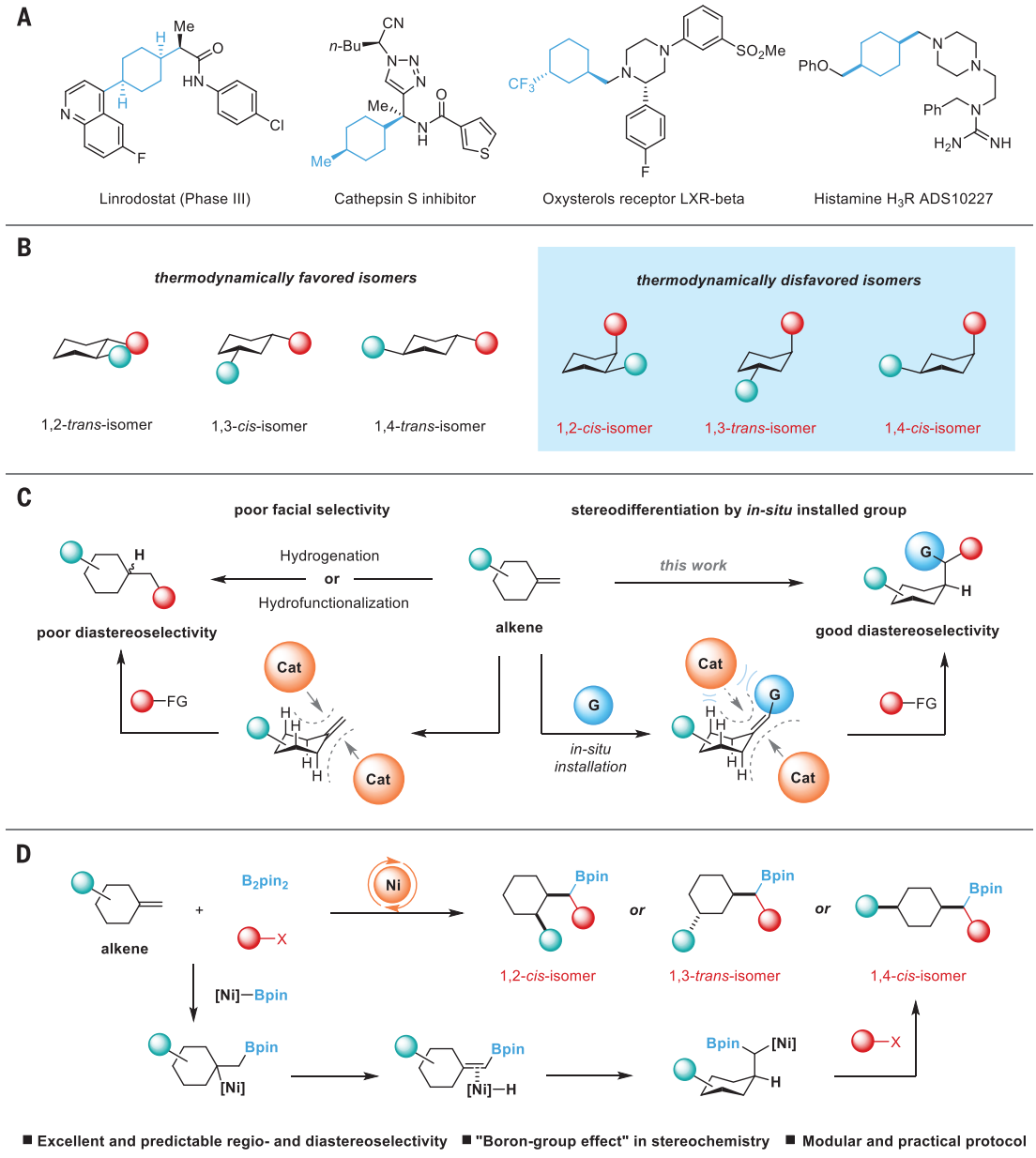


Fig. 1. Synthesis of substituted cyclohexanes with kinetic stereocontrol. (**A**) Representative bioactive molecules containing thermodynamically disfavored disubstituted cyclohexanes. Me, methyl; *n*-Bu, *n*-butyl; Ph, phenyl. (**B**) Thermodynamically favored versus disfavored disubstituted cyclohexanes. (**C**) Previous studies and the conception of this work. FG, functional group; Cat, catalyst; G, group. (**D**) Execution of this work: modular synthesis of substituted cyclohexanes with kinetic stereocontrol via nickel-catalyzed migratory carboboration of substituted methylenecyclohexanes.

produced by the entire [Ni]-H system, favors the installation of the hydrogen atom in the equatorial plane where the Bpin group and PyrOx ligand are placed at longer distances (the same effect is observed for 1,3 insertion; see supplementary materials for details). In contrast, low stereodifferentiation is found in the absence of the Bpin group, with both transition states almost isoenergetic owing to the larger conformational flexibility of the PyrOx ligand in both axial and equatorial positions.

We next focused on the generality of this nickel-catalyzed reaction, first with a series of 4-substituted methylenecyclohexanes (Fig. 3A). The corresponding secondary alkyl boronic esters could be obtained in moderate to good yields with exclusive 1,1-regio- and excellent 1,4-*cis*-stereoselectivity (from 91:9 to >99:1 dr) when a diversity of substituents including alkyl, ester, and aryl groups (**9** to **25** and **29** to **31**), as well as heteroatoms (**26** to **28**), were attached. Notably, even with a small methyl substituent (**32**), good diastereoselectivity (dr = 91:9) could be obtained. Moreover, the resulting cyclohexane frameworks (**20** and **23** to **25**) have widespread applications in liquid crystalline materials (*32*, *33*). Variation of the substituent on the arene ring of the benzylic halide did not affect the efficiency or stereoselectivity. Next, we examined methylenecyclohexanes bearing a 3-substitutent and obtained excellent 1,3-*trans*-diastereoselectivities (Fig. 3B). Furthermore, the 1,2-*cis*-cyclohexane derivatives were obtained from 2-substituted methylenecyclohexanes with the same high diastereoselectivity (Fig. 3C). Besides six-membered carbon rings, the heterocyclic piperidine derivatives (**42** to **44**) were also compatible with the developed protocol, and

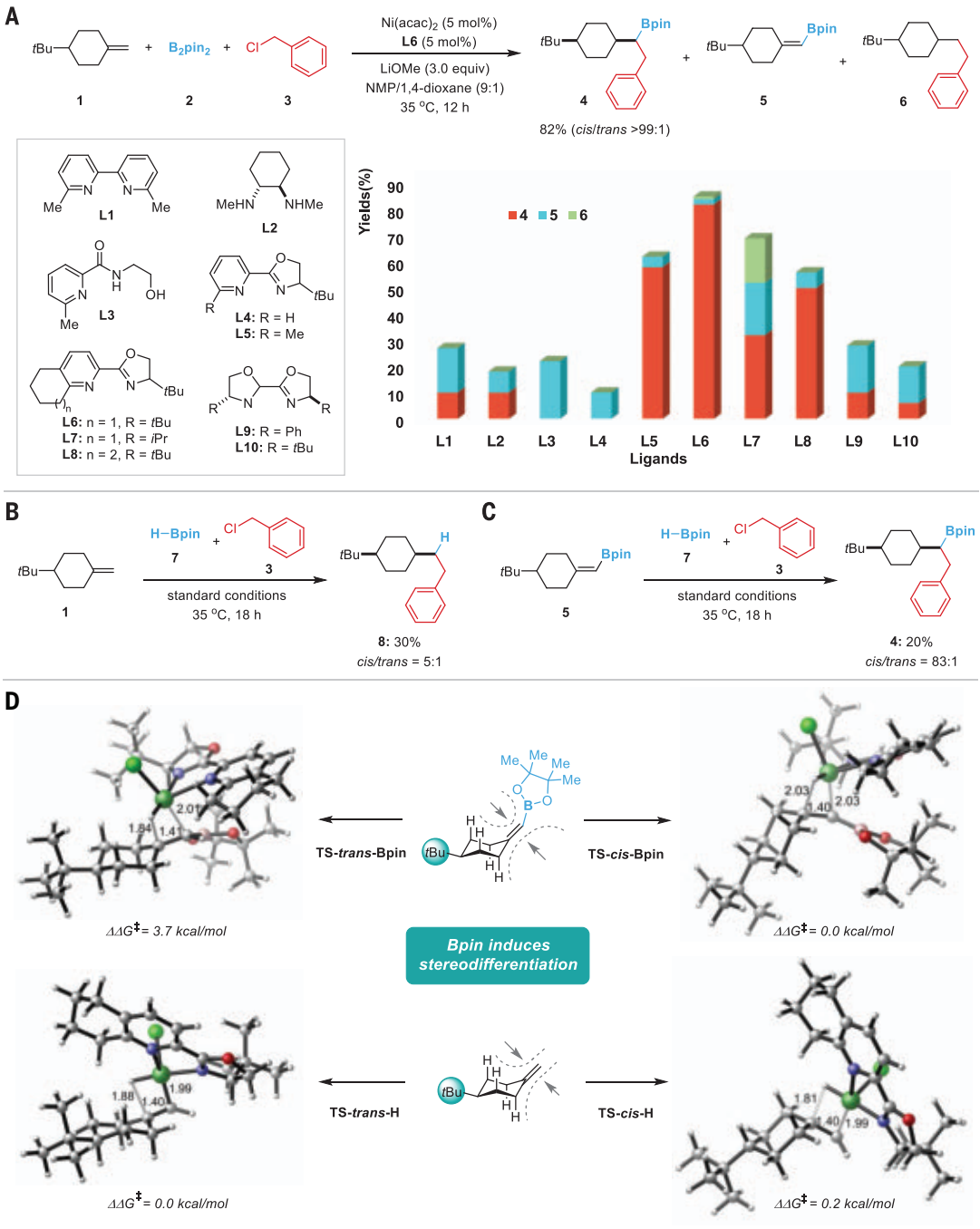


Fig. 2. Reaction development and proof of concept. (**A**) Identification of the optimal reaction conditions. R, radical group; *t*Bu, *tert*-butyl; *i*Pr, isopropyl. (**B**) Hydroalkylation of terminal alkene. (**C**) Hydroalkylation of alkenylboron. (**D**) Density functional theory computational studies on the role of the Bpin group in the stereochemistry of this reaction. ΔΔG‡, activation Gibbs free energy difference referred to the most stable transition state; TS, transition state. The numbers within the ball-and-stick models represent the bond lengths in angstroms.

the corresponding kinetic products were obtained with high facial selectivity (Fig. 3D). The two-component coupling products or alkene isomerization products make up most of the remaining mass balance of these reactions (see supplementary materials for details of representative examples). The stereochemistry (**13**, **16**, **22**, **25**, **27**, **28**, **32**, and **41**) was unambiguously confirmed by single-crystal x-ray crystallography. A wide range of flexible functional groups including aryl bromide (**13**), esters (**16** and **19**), phenol (**21**), cyano (**22**), ketal (**25**), benzylether (**26**), imide (**27**), and amide (**28**) were all well tolerated under the mild reaction conditions.

Encouraged by the above results, we subsequently investigated architecturally complex molecules bearing a multiply substituted methylenecyclohexane moiety. As illustrated in Fig. 3E, the complex molecules derived from octahydroindole (**45**), lithocholic acid (**46**), nootkatone (**47**), (–)-menthone (**48**), and drostanolone propionate (**49**), which bear diverse substitution patterns, were all transformed into the single product isomer by this protocol. In addition, (–)-caryophyllene oxide (**50**) bearing a terminal alkene attached to a nine-membered

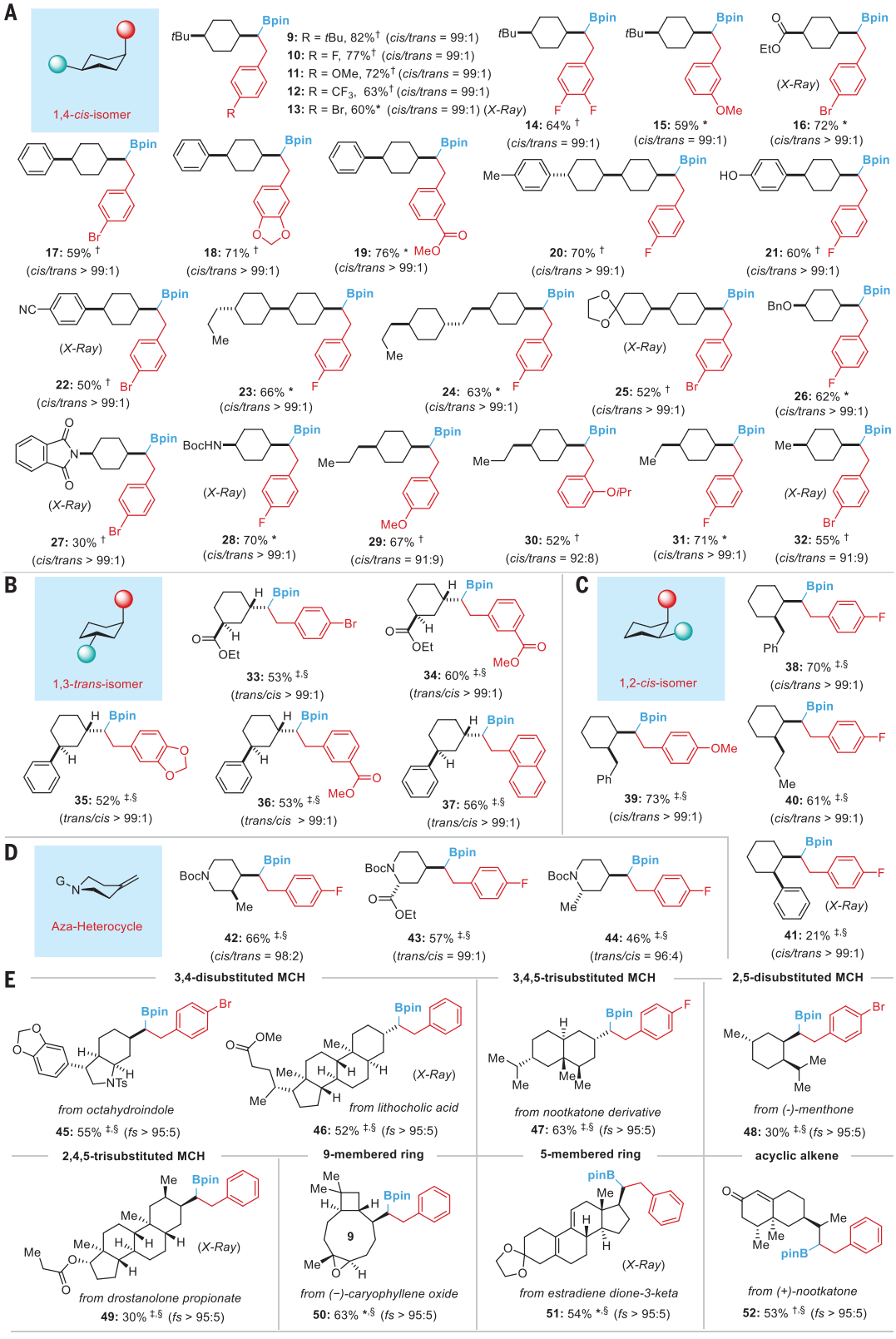

A
1,4-cis-isomer
9: R = tBu, 82%† (cis/trans = 99:1)
10: R = F, 77%† (cis/trans = 99:1)
11: R = OMe, 72%† (cis/trans = 99:1)
12: R = CF3, 63%† (cis/trans = 99:1)
13: R = Br, 60%* (cis/trans = 99:1) (X-Ray)
14: 64% † (cis/trans = 99:1)
15: 59% * (cis/trans = 99:1)
16: 72% * (cis/trans > 99:1)
17: 59% † (cis/trans > 99:1)
18: 71% † (cis/trans > 99:1)
19: 76% * (cis/trans > 99:1)
20: 70% † (cis/trans > 99:1)
21: 60% † (cis/trans > 99:1)
22: 50% † (cis/trans > 99:1)
23: 66% * (cis/trans > 99:1)
24: 63% * (cis/trans > 99:1)
25: 52% † (cis/trans > 99:1)
26: 62% * (cis/trans > 99:1)
27: 30% † (cis/trans > 99:1)
28: 70% * (cis/trans > 99:1)
29: 67% † (cis/trans = 91:9)
30: 52% † (cis/trans = 92:8)
31: 71% * (cis/trans > 99:1)
32: 55% † (cis/trans = 91:9)
B
1,3-trans-isomer
33: 53% ‡,§ (trans/cis > 99:1)
34: 60% ‡,§ (trans/cis > 99:1)
35: 52% ‡,§ (trans/cis > 99:1)
36: 53% ‡,§ (trans/cis > 99:1)
37: 56% ‡,§ (trans/cis > 99:1)
C
1,2-cis-isomer
38: 70% ‡,§ (cis/trans > 99:1)
39: 73% ‡,§ (cis/trans > 99:1)
40: 61% ‡,§ (cis/trans > 99:1)
D
Aza-Heterocycle
42: 66% ‡,§ (cis/trans = 98:2)
43: 57% ‡,§ (trans/cis = 99:1)
44: 46% ‡,§ (trans/cis = 96:4)
41: 21% ‡,§ (cis/trans > 99:1)
E
3,4-disubstituted MCH
3,4,5-trisubstituted MCH
2,5-disubstituted MCH
from octahydroindole
45: 55% ‡,§ (fs > 95:5)
from lithocholic acid
46: 52% ‡,§ (fs > 95:5)
from nootkatone derivative
47: 63% ‡,§ (fs > 95:5)
from (-)-menthone
48: 30% ‡,§ (fs > 95:5)
2,4,5-trisubstituted MCH
9-membered ring
5-membered ring
acyclic alkene
from drostanolone propionate
49: 30% ‡,§ (fs > 95:5)
from (−)-caryophyllene oxide
50: 63% *,§ (fs > 95:5)
from estradiene dione-3-keta
51: 54% *,§ (fs > 95:5)
from (+)-nootkatone
52: 53% †,§ (fs > 95:5)

Fig. 3. Reaction scope. (**A**) Reactions of 4-substituted methylenecyclohexanes. (**B**) Reactions of 3-substituted methylenecyclohexanes. EtO, ethylene oxide; OMe, methoxy group; BocHN, *tert*-butyl carbamate; O*i*Pr, isopropoxy; OEt, ethoxy functional group. (**C**) Reactions of 2-substituted methylenecyclohexanes. (**D**) Reactions of aza-hetereocyclic alkenes. (**E**) Late-stage modification of complex molecules. The asterisk symbol denotes isolated yield of the borylated product. The single-dagger symbol denotes isolated yield of the alcohol after H_2O_2 oxidation. The double-dagger symbol denotes isolated yield of the ketone after Dess-Martin oxidation. The section symbol indicates that no relative stereoselectivity was observed for the carbon center bearing the Bpin group in this case. MCH, methylenecyclohexane.

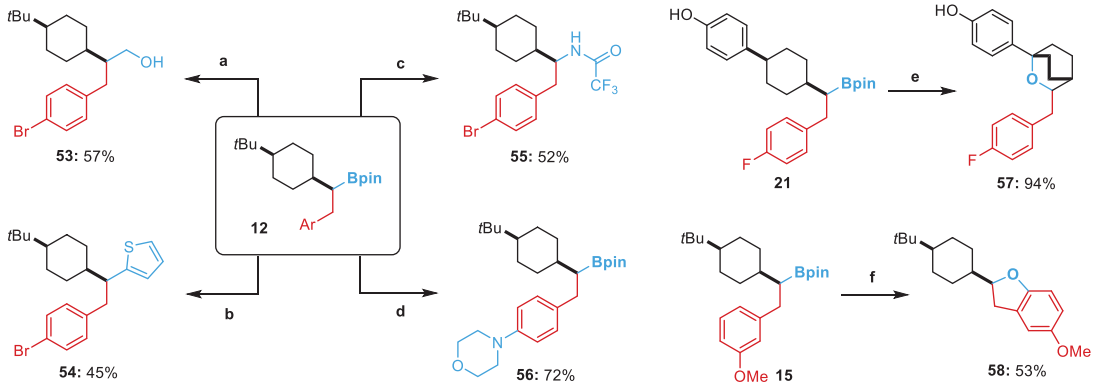


Fig. 4. Diversifications of the representative products.

ring, estradiene dione-3-keta (**51**) bearing a five-membered ring, and (+)-nootkatone (**52**) with an acyclic terminal alkene could also smoothly participate in this nickel-catalyzed reaction to deliver the corresponding products with excellent facial selectivity (fs > 95:5). These results demonstrate the compatibility and potential of this protocol in late-stage diversification of complex bioactive molecules.

The incorporation of boron functional groups into molecules is particularly attractive because of their diverse downstream transformations (*34*). To further illustrate the synthetic potential of this method, additional diversifications of products were conducted. As illustrated in Fig. 4, using *cis*-**13** as the model substrate, the carbon-boron bond could be transformed into C-C(sp^3) (**53**) or C-C(sp^2) (**54**) bonds by Matteson reactions (*35*), as well as a C–N bond by Liu's amination protocol (**55**) (*36*). Moreover, a palladium-catalyzed cross-coupling of the aryl bromide moiety offered the amination product with the boronic ester intact (**56**) (*37*). Consequently, several structurally diverse *cis*-1,4-disubstituted cyclohexanes were obtained. In addition, bicyclic ether **57** and heterobicyclic compound **58** were also readily prepared from the products (*38*, *39*).

To highlight the value of this protocol, a comparative study was performed using classical cross-coupling methodologies and our method (fig. S1). In general, cyclohexanones (or cyclohexanols) are readily available compounds, which can undergo reduction, halogenation, and classical cross-couplings to furnish the thermodynamically favored products (fig. S1, left) (*15*, *40*, *41*). In contrast, the same ketones can undergo olefination followed by this protocol to generate their analogs with a reversed stereocenter (fig. S1, right). What is particularly noteworthy is that this observed stereochemistry not only holds in the simple disubstituted cyclic products (*cis*-**8**) but is also observed for multisubstituted cyclohexane molecules (**60c** and **62c**). This is complementary to current hydrogenation technologies that form the thermodynamically favored products (*18*).

With this work, we have achieved a nickel-catalyzed chemoselective three-component coupling reaction, which provides expedient access to the framework of thermodynamically disfavored substituted cyclohexanes from easily accessible methylenecyclohexanes, B_2pin_2, and benzyl halides. A wide scope of functional groups, including aryl bromide, ester, amine, amide, imide, and α,β-unsaturated ketone, were all compatible with this transformation.

REFERENCES AND NOTES

1. C. M. Marson, *Chem. Soc. Rev.* **40**, 5514–5533 (2011).
2. L. Kiss, F. Fülöp, *Chem. Rev.* **114**, 1116–1169 (2014).
3. M. Aldeghi, S. Malhotra, D. L. Selwood, A. W. E. Chan, *Chem. Biol. Drug Des.* **83**, 450–461 (2014).
4. G. Pattenden, T. V. Lee, in *General and Synthetic Methods: Volume 7*, G. Pattenden, Ed. (The Royal Society of Chemistry, 1985), pp. 310–348.
5. X. Jiang, R. Wang, *Chem. Rev.* **113**, 5515–5546 (2013).
6. J. Jurczyk *et al.*, *Science* **373**, 1004–1012 (2021).
7. E. C. Cherney *et al.*, *ACS Med. Chem. Lett.* **12**, 288–294 (2021).
8. N. Moss *et al.*, *Bioorg. Med. Chem. Lett.* **22**, 7189–7193 (2012).
9. D. A. Claremon *et al.*, European Patent WO2016022521A1·2016-02-11 (2016).
10. M. Staszewski *et al.*, *ACS Chem. Neurosci.* **12**, 2503–2519 (2021).
11. Y. Iida, R. Sugawara, *Agric. Biol. Chem.* **45**, 1553–1559 (1981).
12. S. Reymond, J. Cossy, *Chem. Rev.* **108**, 5359–5406 (2008).
13. H. Du, K. Ding, in *Handbook of Cyclization Reactions: Volume 1*, S. Ma, Ed. (Wiley, 2010), pp. 1–57.
14. G. Masson, C. Lalli, M. Benohoud, G. Dagousset, *Chem. Soc. Rev.* **42**, 902–923 (2013).
15. T. Thaler *et al.*, *Nat. Chem.* **2**, 125–130 (2010).
16. X. Mu, Y. Shibata, Y. Makida, G. C. Fu, *Angew. Chem. Int. Ed.* **56**, 5821–5824 (2017).
17. J. Li, Q. Ren, X. Cheng, K. Karaghiosoff, P. Knochel, *J. Am. Chem. Soc.* **141**, 18127–18135 (2019).
18. M. Cushman *et al.*, *J. Med. Chem.* **37**, 3040–3050 (1994).
19. D. Gärtner, A. Welther, B. R. Rad, R. Wolf, A. Jacobi von Wangelin, *Angew. Chem. Int. Ed.* **53**, 3722–3726 (2014).
20. K. Iwasaki, K. K. Wan, A. Oppedisano, S. W. M. Crossley, R. A. Shenvi, *J. Am. Chem. Soc.* **136**, 1300–1303 (2014).
21. B. K. Peters *et al.*, *J. Am. Chem. Soc.* **138**, 11930–11935 (2016).
22. L. N. Mendelsohn *et al.*, *ACS Catal.* **11**, 1351–1360 (2021).
23. S. A. Green, S. Vásquez-Céspedes, R. A. Shenvi, *J. Am. Chem. Soc.* **140**, 11317–11324 (2018).
24. M. P. Wiesenfeldt, Z. Nairoukh, W. Li, F. Glorius, *Science* **357**, 908–912 (2017).
25. Z. Nairoukh, M. Wollenburg, C. Schlepphorst, K. Bergander, F. Glorius, *Nat. Chem.* **11**, 264–270 (2019).
26. X. Zhang, L. Ling, M. Luo, X. Zeng, *Angew. Chem. Int. Ed.* **58**, 16785–16789 (2019).
27. J.-C. Richer, *J. Org. Chem.* **30**, 324–325 (1965).
28. Y. Li *et al.*, *Angew. Chem. Int. Ed.* **58**, 8872–8876 (2019).
29. W. Wang, C. Ding, G. Yin, *Nat. Catal.* **3**, 951–958 (2020).
30. C. Ding, Y. Ren, C. Sun, J. Long, G. Yin, *J. Am. Chem. Soc.* **143**, 20027–20034 (2021).
31. Y. Zhang, X. Xu, S. Zhu, *Nat. Commun.* **10**, 1752 (2019).
32. M. Hird, *Chem. Soc. Rev.* **36**, 2070–2095 (2007).
33. Y. Fujisawa, A. Asano, Y. Itoh, T. Aida, *J. Am. Chem. Soc.* **143**, 15279–15285 (2021).
34. C. Sandford, V. K. Aggarwal, *Chem. Commun.* **53**, 5481–5494 (2017).
35. D. S. Matteson, *Acc. Chem. Res.* **21**, 294–300 (1988).
36. X. Liu *et al.*, *Angew. Chem. Int. Ed.* **59**, 2745–2749 (2020).
37. J. C. Lee, M. G. Wang, F. E. Hong, *Eur. J. Inorg. Chem.* **2005**, 5011–5017 (2005).
38. A. M. Hanson *et al.*, *J. Med. Chem.* **61**, 4720–4738 (2018).
39. J. Alvarado, J. Fournier, A. Zakarian, *Angew. Chem. Int. Ed.* **55**, 11625–11628 (2016).
40. X. Yu, T. Yang, S. Wang, H. Xu, H. Gong, *Org. Lett.* **13**, 2138–2141 (2011).
41. C. Zhao, X. Jia, X. Wang, H. Gong, *J. Am. Chem. Soc.* **136**, 17645–17651 (2014).

ACKNOWLEDGMENTS

We thank G. Liu at the Shanghai Institute of Organic Chemistry and F. Schoenebeck at RWTH Aachen University for their insightful discussion. **Funding:** Funding was provided by National Natural Science Foundation of China grants 21871211 and 22122107 (G.Y.) and Fundamental Research Funds for Central Universities 2042021kf0190 (G.Y.). High Performance Computing Centre "Beronia" provided computational resources (I.F.-A.). The University of La Rioja provided financial support (I.F.-A.). **Author contributions:** Conceptualization: G.Y. Methodology: All authors. Investigation: Ya.L., H.S., H.W., H.L., and I.F.-A. Visualization: I.F.-A. and G.Y. Funding acquisition: I.F.-A. and G.Y. Project administration: I.F.-A. and G.Y. Supervision: G.Y. Writing – original draft: Ya.L., I.F.-A., and G.Y. Writing – review & editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Solid-state structures of the following compounds are freely available from the Cambridge Crystallographic Data Centre under the associated CCDC codes: **13**, 2159963; **16**, 2159966; **22b**, 2159971; **25b**, 2159970; **27b**, 2161229; **28**, 2160663; **32b**, 2159973; **41c**, 2159972; **46c**, 2159976; **49c**, 2160636; **51**, 2160345; and **62**, 2159981.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn9124
Materials and Methods
Figs. S1 to S6
Tables S1 to S5
NMR Spectra
References (*42*–*49*)

Submitted 17 January 2022; accepted 6 April 2022
10.1126/science.abn9124

BIOMECHANICS

Recovery mechanisms in the dragonfly righting reflex

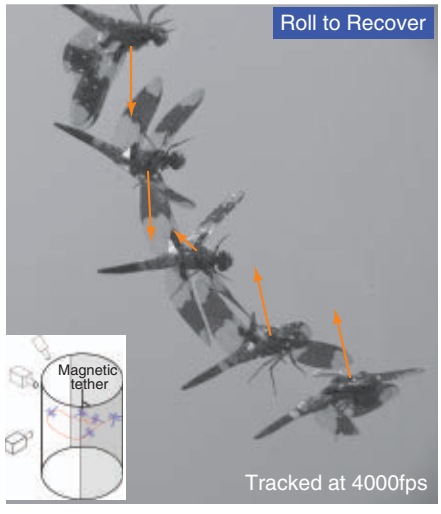
Z. Jane Wang[1,2,3]*, James Melfi Jr.[2], Anthony Leonardo[4]

Insects have evolved sophisticated reflexes to right themselves in mid-air. Their recovery mechanisms involve complex interactions among the physical senses, muscles, body, and wings, and they must obey the laws of flight. We sought to understand the key mechanisms involved in dragonfly righting reflexes and to develop physics-based models for understanding the control strategies of flight maneuvers. Using kinematic analyses, physical modeling, and three-dimensional flight simulations, we found that a dragonfly uses left-right wing pitch asymmetry to roll its body 180 degrees to recover from falling upside down in ~200 milliseconds. Experiments of dragonflies with blocked vision further revealed that this rolling maneuver is initiated by their ocelli and compound eyes. These results suggest a pathway from the dragonfly's visual system to the muscles regulating wing pitch that underly the recovery. The methods developed here offer quantitative tools for inferring insects' internal actions from their acrobatics, and are applicable to a broad class of natural and robotic flying systems.

Dragonflies are some of the most ancient insects, and they offer important clues to the evolution of flight control strategies. Among their acrobatic feats is an ability to right themselves when falling upside down (*1*). Aerial righting behavior has been found in different animal taxa, including small mammals (*2*), flightless insects (*3*), reptiles (*4*), flying insects (*5*), birds (*6*), and bats (*7*), suggesting a convergent behavior that may shed light on the origin of flight (*8*). There has been much progress in the experimental characterization of aerial maneuvers (*5*, *9*–*17*). In particular, high-speed video tracking has offered increasingly detailed flight kinematics since Marey's classical films of a falling cat. Obtaining well-resolved three-dimensional (3D) flight kinematics is, however, nontrivial; careful experiments and calibrations are required. Understanding the physical mechanisms of righting further requires quantitative models of flight (*18*–*20*). Here, we developed experimental and computational methods for analyzing aerial maneuvers, including free-flight simulations that account for both unsteady aerodynamics and body-wing inertial coupling, so as to tease out the key wing asymmetries that lead to the exhibited maneuver. These subtle wing modulations provide important clues to the internal control schemes for aerial maneuvers. Dragonflies control each wing separately with direct muscles. Our computations, together with statistical analyses, show that the observed rolling motion results primarily from wing pitch asymmetry. Our behavioral experiments further show that vision is critical for eliciting the observed rolling response. These results imply pathways between the dragonfly's visual system and the direct muscles that modulate the wing pitch (Fig. 1).

To start, we set up an experiment in which the dragonflies were released from a magnetic tether from different orientations (Fig. 1 and movie S1). The dragonflies performed stereotypical maneuvers (*1*), which depend on their initial body orientation (fig. S3), offering the possibility for controlled studies. To understand righting mechanisms in response to free fall, it is necessary to design an experiment in which the righting reflex can be differentiated from other reflexes such as those involved in take-off or loss of leg contact with the perch. For example, a rapid retraction of a perch away from the dragonfly (*21*), to induce a fall, elicits a reflex in the legs as well as a subsequent righting reflex; the two are interleaved in time and are difficult to discriminate. It is also critical to have reproducible initial starting conditions and to describe the dynamic states with proper definition of body translation and rotation. Thus, we focused on experiments in which the legs were free from contact during the initial release to avoid a confounding tarsal reflex, and we analyzed the recovery maneuvers from falling upside down with nearly identical initial conditions (supplementary text).

When falling upside down from a resting state, dragonflies rolled 180° to reorient. The righting episode lasted for about 200 ms (7 dragonflies, 37 trials, 224 ± 34 ms), which includes a reaction time of 99 ± 26 ms and active recovery of 125 ± 28 ms (three to five wing beats). The body rolled at a nearly steady rate, with a sharp transition at the beginning and the end of the episode (Fig. 2). This repeatable maneuver allowed us to quantify flight patterns in detail. To obtain the time series of the wing and body kinematics at high spatial and temporal resolutions, we filmed flights using three high-speed video cameras at 4000 fps,



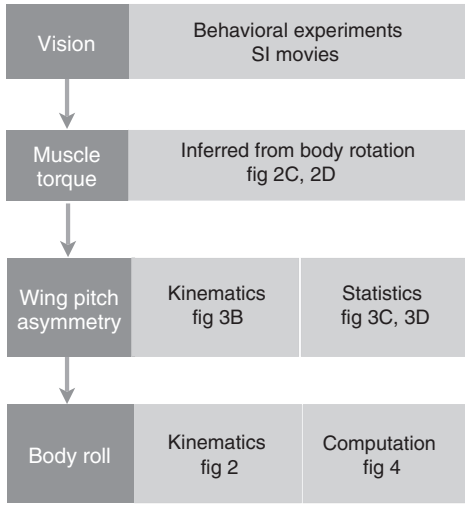


Fig. 1. The key actions along the recovery pathway and the corresponding analysis methods. The dragonflies were released from a magnetic tether in an upside-down orientation and were filmed using three high-speed cameras at 4000 fps. The dragonflies rolled 180° to reorient themselves by means of left and right wing pitch asymmetry, which was elicited by two visual systems: the ocelli and the compound eyes. To tease out the wing pitch asymmetry, we used computational simulation, based on the 3D kinematic data, in addition to statistical analyses. To deduce the driving torque, we constructed a method based on the Euler equations applied to the body rotation. To identify the initial response, we carried out behavioral experiments with the dragonfly's vision altered. These analyses revealed a pathway from the dragonfly's vision to body rotation during its recovery maneuver over ~200 ms.

[1]Department of Physics, Cornell University, Ithaca, NY 14850, USA. [2]Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14850, USA. [3]Center for Computational Biology, Flatiron Institute, New York, NY 10010, USA. [4]Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA.
*Corresponding author. Email: jane.wang@cornell.edu

so that each wing beat (30 to 40 Hz) contained about 100 points of measurement to resolve the sub–wing-beat kinematics. All four wings and the body were tracked. We used a semiautomated tracking algorithm (*22*) and monitored each frame to ensure that the markers were tracked well, and selected flights with all markers visible for full analyses. We then reprojected the tracked markers back to the original film to double-check the conversions and the calibrations (movie S1). The reconstructed 3D kinematics from the markers are represented in the time series of 15 Euler angles and three body positions at the estimated center of mass (Figs. 2 and 3 and figs. S4 to S10).

Measured wing motions showed a complex mixture of modulations in stroke angle, phase shift, inclination of the stroke plane, and wing pitch orientation, with variations from trial to trial. To detect the relevant shifts in wing motion, we first calculated the statistical correlation between different wing variables and body angular velocity, and found the highest correlation associated with wing pitch asymmetry (Fig. 3C). This was further checked by grouping the flights according to whether the body rolled in a clockwise or counterclockwise direction. The reversal of pitch asymmetry was indeed correlated with the reversal of body rotation (Fig. 3D). These correlations suggest that the wing pitch asymmetry between the left and right wings led to the body roll. Still, they do not rule out the contributions of other asymmetries. In particular, wing stroke angles also exhibited apparent shifts during recovery (Fig. 3B), which warrants further study.

The relationship between wing motion and body motion is nonlinear; it is determined by the Euler equations for body rotation, aerodynamic forces, and the coupling among the inertial and aerodynamic effects. Each wing asymmetry can lead to a rotational mode that has a mixture of roll, yaw, and pitch. To understand the causal relationship between wing asymmetry and body rotation, we developed a flight model of four-winged insects (see supplementary materials). The model simulates the flight in six degrees of freedom, taking into account both unsteady aerodynamic forces (*18*) and instantaneous inertial coupling between wings and body motions (*23*), and was tested on hovering and maneuvering flight (figs. S13 and S14). Using this flight model, we compared the effects of separately applying wing pitch or wing stroke asymmetries, the two leading candidates for body roll (Fig. 3C). Computations showed that these two asymmetries elicited markedly different body rotations (Fig. 4). Wing pitch asymmetry induced predominantly a body roll (Fig. 4A and movie S2), confirming the observed statistical correlation. Moreover, when the amount of wing asymmetry was varied, the resulting changes in body rotation followed proportionally, consistent with experimental data (fig. S18). In contrast, stroke amplitude asymmetry induced predominantly a yaw motion of the body (Fig. 4B). The roll motion was substantially smaller and insufficient to return the dragonfly to its normal orientation (movie S3). These results further show that dragonflies rely on a wing pitch asymmetry between the left and right wings to drive roll recovery. We further note that the averaged motion of the left and right wings during the recovery flight accelerates the dragonfly forward, and is thus different from the typical hovering wing motion that would have accelerated the dragonfly downward, given its initial upside-down orientation. It is this subtle interplay between wing asymmetry and average wing motion that leads to the difference in dragonflies' and flies' maneuvers, further shown computationally (fig. S18).

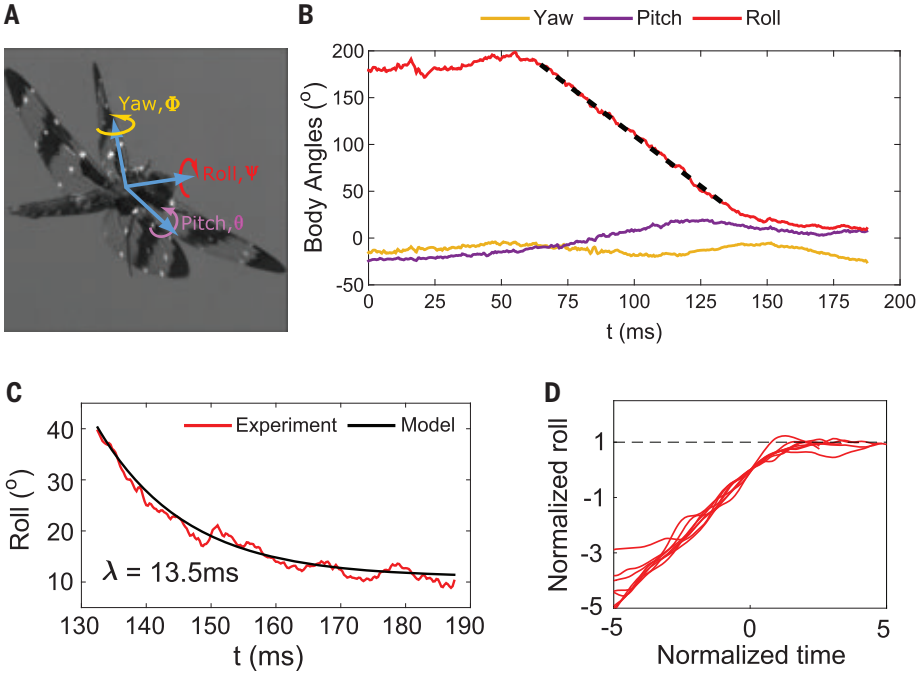


Fig. 2. Body kinematics and a method for estimating the driving torque. (**A**) Definition of the body Euler angles. (**B**) Measured body Euler angles during one flight reveal the dominant rolling motion during the recovery sequence. (**C**) Zoomed-in view of the rolling angle at the end of the rolling motion. The analysis of the shape of this transitional curve leads to an estimate for the driving torque. (**D**) The roll angle from different recovery sequences during the transition follows a single curve after a rescaling in time and angle, consistent with the model.

To infer the magnitude of the driving torque, we constructed a quick method that can be applied to the experimental data without resorting to computations. The trajectory of the roll angle contains a linear segment followed by a sharp transition (Fig. 2B). Because the roll dominates, the Euler equation for the roll angular rate is simplified to $I\dot{\Omega} \sim -\beta\Omega + \tau$, where τ is the driving torque and $-\beta\Omega$ is the effective damping torque due to aerodynamic drag induced by the sum of symmetric wing motion and body rotation (*24*). Angular velocity is relatively constant during recovery, $\dot{\Omega} \sim 0$; therefore, $\tau = \beta\Omega$. To find β, we note that at the tail end of the roll, the dynamics are governed by $I\dot{\Omega} \sim -\beta\Omega$, or $\Omega = \Omega_0[\exp(-t/\lambda)]$, where $\lambda = I/\beta$ is the time constant. Therefore, the roll angle is given by $\Psi(t) - \Psi(t_0) = \Omega_0\lambda[1 - \exp(-t/\lambda)]$, where λ can be estimated from the data. In the case shown, $\lambda \sim 13.5$ ms, which is roughly half a wing beat. This, together with body inertia, gives an estimate of the torque that rotates the dragonfly, $\tau \sim \beta\Omega_0 = I\Omega_0/\lambda \sim 21\,\mu$N·m. Comparing it to the nominal torque on each wing during steady state, $\tau_0 = MgL/4$, gives a relative scale. For a typical body weight M of 300 mg, with the net force acting at $L \sim 20$ mm away from the wing root, the driving torque requires an increase or decrease of one-third of the nominal torque on each wing.

The wing symmetry needed for maneuvers is elicited from upstream sensory signals. How does a dragonfly perceive its fall? Insects are equipped with many sensors to detect their orientation and movement in space (*25*). In general, they use multiple feedback systems to ensure robust and versatile flight control (*25*–*27*). Possible candidates include their antenna, visual systems, legs, and other mechanosensors. We investigated which of these drives the righting reflex. Dragonflies, like other insects, have

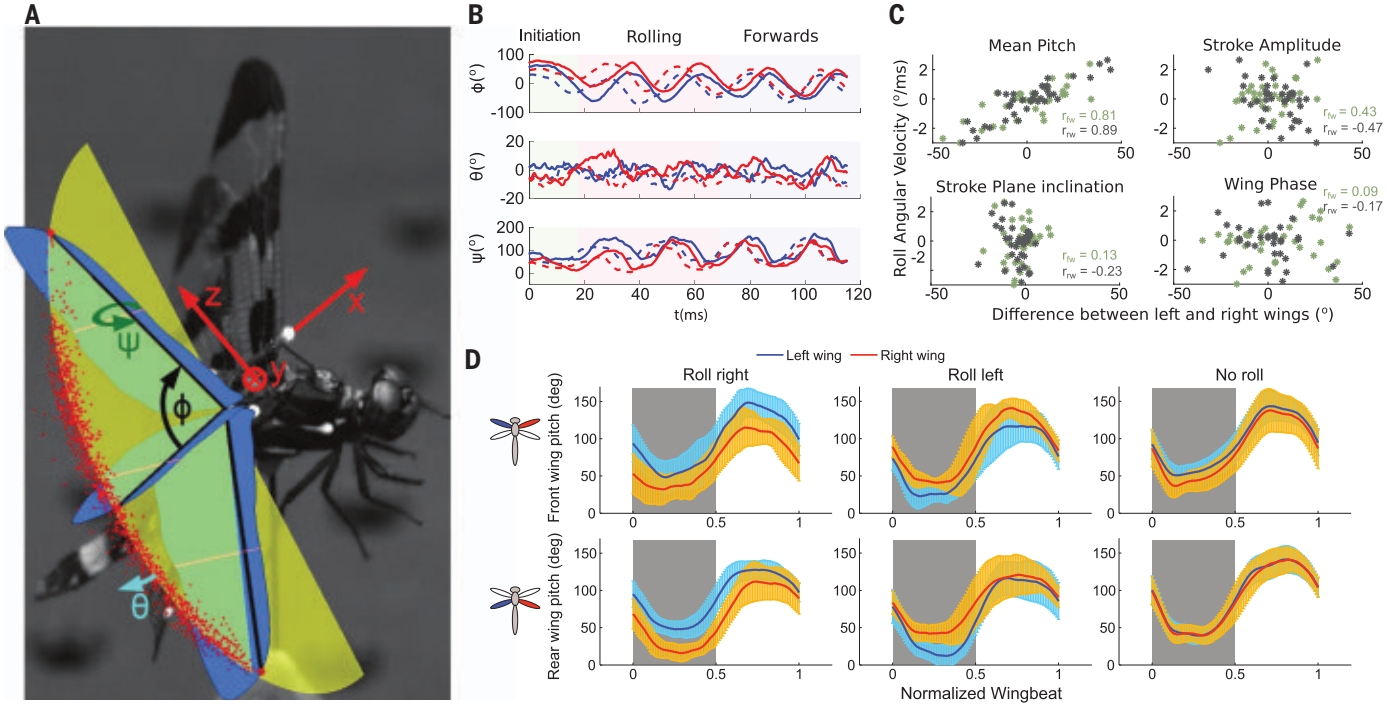


Fig. 3. Wing kinematics and statistics show a strong correlation between wing pitch asymmetry and body roll angular velocity. (**A**) The definition of the wing Euler angles: the stroke angle (ϕ), the wing pitch angle relative to the inclined stroke plane (ψ), and the deviation angle of the leading edge from the stroke plane (θ). (**B**) The measured wing Euler angles during one flight. Additional kinematics are shown in figs. S4 to S10. Blue, left wing; red, right wing; solid line, forewing; dashed line, hindwing. (**C**) The correlation between the body roll angular velocity and the left-right wing asymmetries from multiple experiments. Each data point represents the averaged value over a wing beat for the front (blue) and rear (red) pair of wings (5 dragonflies, 55 wing beats). (**D**) Details of the wing pitch during a wing beat in three scenarios: rolling right, rolling left, and forward flight. The solid line represents the averaged value; the colored region denotes one standard deviation. Top panels, front wings; bottom panels, rear wings. The reverse in wing pitch asymmetry leads to a reverse in the rolling direction, confirming the strong correlation between the two.

two visual systems: the compound eyes and the ocelli. The ocelli are three simple eyes that can serve as a horizon detector and trigger a head rotation toward the light (*28*, *29*) We performed behavioral experiments on dragonflies with blocked vision to investigate the role of vision in the righting reflex. The reaction time statistics show that vision is critical for eliciting the observed rolling behavior. When their visual systems were intact, dragonflies reacted in ~100 ms and had recovered normal orientation in ~220 ms (7 dragonflies, 37 trials, reaction time 99 ± 26 ms) (movie S1). When the ocelli were blocked, the reaction time became longer (3 dragonflies, 13 trials, reaction time 158 ± 93 ms), and all 13 recovery attempts were unsuccessful to gain level flight within the filming volume (movie S4). When the compound eyes were blocked, the recovery flight was more erratic, occasionally with the head plunging down first, followed by a body pitch assisted by the abdominal bending motion (2 dragonflies, 14 trials, reaction time 156 ± 63 ms, total roll recovery time 272 ± 35 ms) (movie S5). When both the ocelli and the compound eyes were blocked, dragonflies showed a mixture of behavior. In many cases, they did not flap their wings, falling like leaves within the filming volume (3 dragonflies, 35 trials) (movie S6). In the cases where they flapped wings, none of the flapping motion elicited the rolling response (3 dragonflies, 13 trials) (S7). The wings typically flapped after the body passively pitched, or just before reaching the ground (movie S6). Although these experiments do not rule out the participation of other mechanosensory pathways during the recovery maneuver, they suggest that the visual system is the first and dominant response pathway, and the combination of ocelli and compound eyes ensures an accurate recovery. Related work (*30*) using a setup similar to ours (*1*) also found that vision plays a main role in hoverflies' aerial righting.

Dragonflies often perform mid-air righting maneuvers during their prey-capture flights (*15*) and must recover their normal position from different orientations. To right from an upside-down orientation, a dragonfly can either roll or pitch 180°; we have found that it almost always does so using the energetically less costly roll maneuver. Moreover, dragonflies of different species show the same rolling dynamics, which suggests that this is a generic behavior (supplementary text). When released from different initial orientations, the righting mode transitioned from rolling to pitching motion (fig. S3). In more than 300 experiments under normal vision, dragonflies always actively flapped their wings to right themselves. The passive mechanism suggested in a related study (*21*) refers to the body pitching motion of anesthetized dragonflies with motionless wings, released from a perch with an initial pitch velocity (supplementary text). In awake dragonflies, passive body pitch can occur, before wing flapping, as a result of either an initial leg kick or a back-pitched orientation, but is insufficient for righting. The subsequent flapping motion does not necessarily explore the passive body rotation. It can produce a rolling motion when the body is back-pitched (movie S8) or can pitch the body in the opposite direction of the initial pitch (movie S9). These specific cases, as well as the general maneuvers that involve a mixture of roll, yaw, and pitch, show unequivocally that dragonflies use active mechanisms to right themselves.

Our results point to a pathway from the dragonfly's visual system to the direct muscles regulating wing pitch that drive body roll when

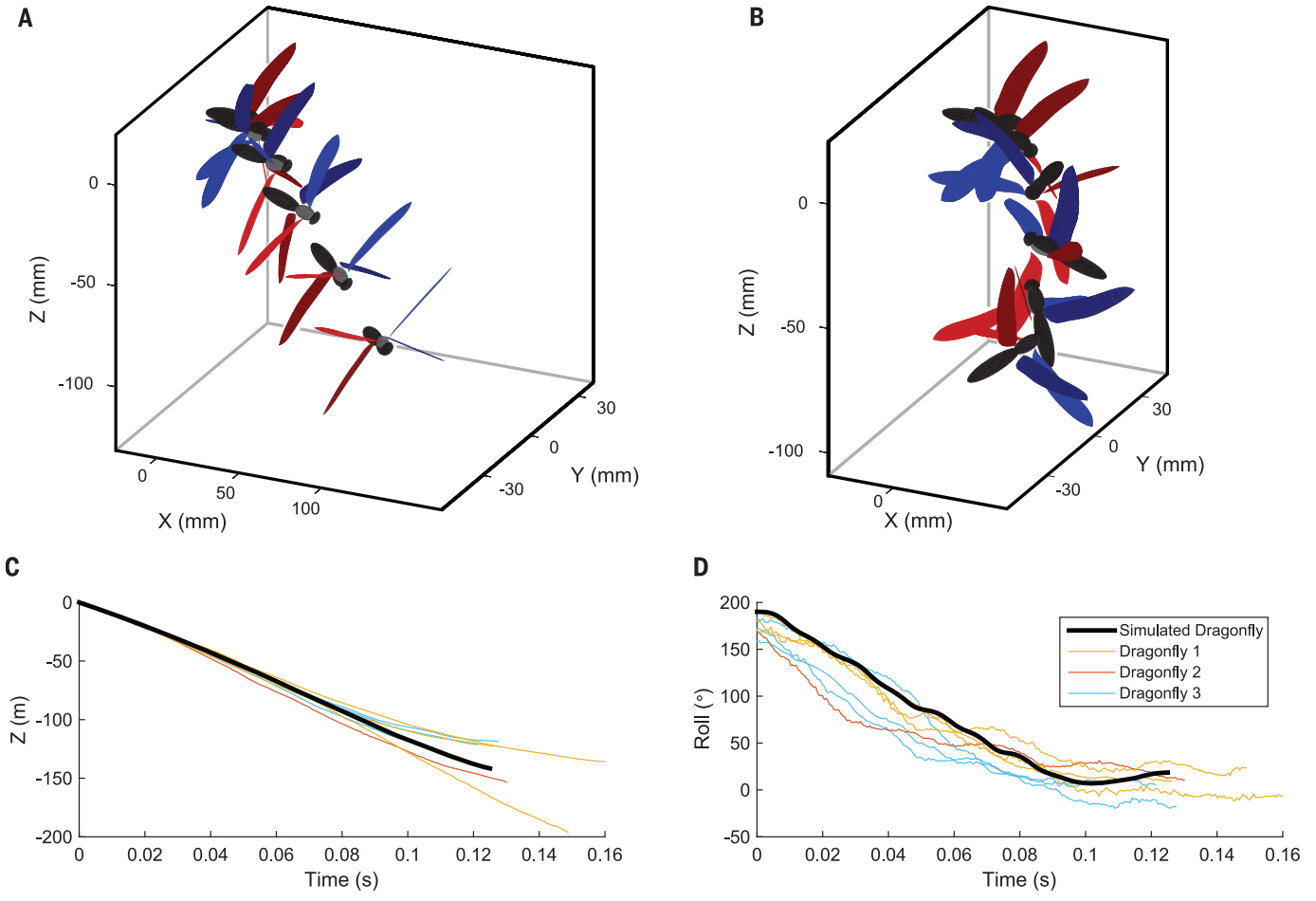


Fig. 4. Computational model compares the effect of wing pitch and wing stroke asymmetries on body rotation. (**A**, **C**, and **D**) Wing pitch asymmetry leads to roll recovery, as seen in the experiments. (**B**) Wing stroke asymmetry, by contrast, leads to a body rotation with a complex mixture of roll, yaw, and pitch and underpredicts the roll seen in experiments. Simulation methods and additional tests are in the supplementary materials.

falling upside down, and imply connections between the early visual neurons and the motor neurons to these muscles. The methods and analyses developed here are not limited to dragonflies. Whereas mammals rely primarily on inertia to right themselves, flying animals use aerodynamic torques in addition to inertia. The computational model developed here can be applied to other insects, birds, and bats (*7*, *12*, *14*–*17*, *27*) and robotic flyers (*31*, *32*). More broadly, our analyses of flight maneuvers provide a noninvasive method for inferring part of the internal control schemes during free flight. In this approach, the physics of flight provides a starting framework for mechanistic explanations of animals' evolved behavior (*20*).

REFERENCES AND NOTES

1. J. Melfi, A. Leonardo, Z. J. Wang, Mid-air recovery of falling dragonflies. APS Division of Fluid Dynamics Gallery of Fluid Motion (2014); https://gfm.aps.org/meetings/dfd-2014/54175ce169702d585c700300.
2. E. J. Marey, *C. R. Hebd. Seances Acad. Sci.* **119**, 714 (1894).
3. S. P. Yanoviak, R. Dudley, M. Kaspari, *Nature* **433**, 624–626 (2005).
4. A. Jusufi, D. I. Goldman, S. Revzen, R. J. Full, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4215–4219 (2008).
5. L. Ristroph *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4820–4824 (2010).
6. D. R. Warrick, M. W. Bundle, K. P. Dial, *Integr. Comp. Biol.* **42**, 141–148 (2002).
7. A. J. Bergou *et al.*, *PLOS Biol.* **13**, e1002297 (2015).
8. R. Dudley, S. P. Yanoviak, *Integr. Comp. Biol.* **51**, 926–936 (2011).
9. S. N. Fry, R. Sayaman, M. H. Dickinson, *Science* **300**, 495–498 (2003).
10. G. Card, M. H. Dickinson, *Curr. Biol.* **18**, 1300–1307 (2008).
11. L. Ristroph, G. J. Berman, A. J. Bergou, Z. J. Wang, I. Cohen, *J. Exp. Biol.* **212**, 1324–1335 (2009).
12. A. J. Bergou, L. Ristroph, J. Guckenheimer, I. Cohen, Z. J. Wang, *Phys. Rev. Lett.* **104**, 148101 (2010).
13. B. Cheng, X. Deng, T. L. Hedrick, *J. Exp. Biol.* **214**, 4092–4106 (2011).
14. F. T. Muijres, M. J. Elzinga, J. M. Melis, M. H. Dickinson, *Science* **344**, 172–177 (2014).
15. M. Mischiati *et al.*, *Nature* **517**, 333–338 (2015).
16. S. Balebail, S. K. Raja, S. P. Sane, *PLOS ONE* **14**, e0219861 (2019).
17. P. Liu, S. Sane, J. Mongeau, J. Zhao, B. Cheng, *Sci. Adv.* **5**, eaax1877 (2019).
18. Z. J. Wang, *Annu. Rev. Fluid Mech.* **37**, 183–210 (2005).
19. M. Sun, *Rev. Mod. Phys.* **86**, 615–646 (2014).
20. Z. J. Wang, *Annu. Rev. Condens. Matter Phys.* **7**, 281–300 (2016).
21. S. T. Fabian, R. Zhou, H. T. Lin, *Proc. R. Soc. B* **288**, 20202676 (2021).
22. T. L. Hedrick, *Bioinspir. Biomim.* **3**, 034001 (2008).
23. S. Chang, Z. J. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11246–11251 (2014).
24. T. L. Hedrick, B. Cheng, X. Deng, *Science* **324**, 252–255 (2009).
25. G. Taylor, H. Krapp, *Adv. Insect Physiol.* **34**, 231–316 (2007).
26. A. Sherman, M. H. Dickinson, *J. Exp. Biol.* **206**, 295–302 (2003).
27. M. H. Dickinson, F. Muijres, *Philos. Trans. R. Soc. B* **371**, rstb.2015.0388 (2016).
28. H. Mittelstaedt, *J. Comp. Physiol. A* **32**, 422–463 (1950).
29. G. Stange, J. Howard, *J. Exp. Biol.* **83**, 351–355 (1979).
30. R. Goulard, J. L. Vercher, S. Viollet, *J. Exp. Biol.* **219**, 2497–2503 (2016).
31. K. Y. Ma, P. Chirarattananon, S. B. Fuller, R. J. Wood, *Science* **340**, 603–607 (2013).
32. Y. Chen *et al.*, *Nature* **575**, 324–329 (2019).

ACKNOWLEDGMENTS

We thank I. Siwanowicz, M. Mischiati, and R. Noest for helpful discussions and H. T. Lin for technical assistance with earlier experiments at Janelia. Z.J.W. is grateful to L. Ristroph, the NYU Applied Mathematics Lab, M. Shelley, the Flatiron Institute, C. Sbrocco, K. Li, and S. Childress for help and support during 2021 summer experiments. **Funding:** Supported by the Visiting Scientist Program, Janelia Research Campus/HHMI (Z.J.W. and A.L.),

and the Simons Fellowship in Mathematics (Z.J.W.). **Author contributions:** Z.J.W. and A.L. designed the research. J.M., Z.J.W., and A.L. carried out experiments at Janelia/HHMI, and Z.J.W. carried out experiments at AML/NYU. Z.J.W. developed the computational and analysis methods. J.M. carried out computations. Z.J.W., J.M., and A.L. analyzed the data. Z.J.W., J.M., and A.L. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the study are present in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg0946
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S4
Movies S1 to S9
References (*33*–*45*)

Submitted 16 December 2020; accepted 5 April 2022
10.1126/science.abg0946

FOREST ECOLOGY

Cross-biome synthesis of source versus sink limits to tree growth

Antoine Cabon[1]*, Steven A. Kannenberg[1], Altaf Arain[2,3], Flurin Babst[4,5], Dennis Baldocchi[6], Soumaya Belmecheri[5], Nicolas Delpierre[7,8], Rossella Guerrieri[9], Justin T. Maxwell[10], Shawn McKenzie[2,3], Frederick C. Meinzer[11], David J. P. Moore[4], Christoforos Pappas[12,13], Adrian V. Rocha[14], Paul Szejner[15], Masahito Ueyama[16], Danielle Ulrich[17], Caroline Vincke[18], Steven L. Voelker[19], Jingshu Wei[20], David Woodruff[11], William R. L. Anderegg[1]

Uncertainties surrounding tree carbon allocation to growth are a major limitation to projections of forest carbon sequestration and response to climate change. The prevalence and extent to which carbon assimilation (source) or cambial activity (sink) mediate wood production are fundamentally important and remain elusive. We quantified source-sink relations across biomes by combining eddy-covariance gross primary production with extensive on-site and regional tree ring observations. We found widespread temporal decoupling between carbon assimilation and tree growth, underpinned by contrasting climatic sensitivities of these two processes. Substantial differences in assimilation-growth decoupling between angiosperms and gymnosperms were determined, as well as stronger decoupling with canopy closure, aridity, and decreasing temperatures. Our results reveal pervasive sink control over tree growth that is likely to be increasingly prominent under global climate change.

Forest ecosystems currently constitute a net carbon (C) sink that offsets ~25% of yearly anthropogenic C emissions, thus actively mitigating climate change (*1*). C allocation to aboveground wood biomass is the largest contributor to vegetation C storage over climate-relevant time scales. However, wood C allocation is poorly understood and is a major uncertainty for projections of future forests' C storage potential (*2*). The common representation of wood growth as a linear function of C assimilation has been identified as a major structural limitation of current vegetation models (*3*, *4*). The development of improved C allocation schemes currently lacks a solid empirical and mechanistic basis (*5*). Thus, there is an urgent need to illuminate the relationship between C assimilation and tree growth.

A fundamental debate revolves around the degree to which C assimilation via photosynthesis (source limitation) versus direct environmental limitations to cambial cell development (sink limitation) controls wood growth (*6*). As reflected by C allocation schemes in the vast majority of vegetation models, source limitation has been the dominant paradigm for decades (*4*). Yet a growing body of literature indicates that cambial activity is typically more sensitive than photosynthesis to a range of environmental conditions, including low water availability, temperature, and nutrient availability (*7*–*11*). The prevalence of source versus sink limitations to tree growth has far-reaching implications for forest dynamics under climate change, because these processes will likely respond differently to global change (*6*–*9*), potentially shifting C allocation away from the stem. Substantial indirect evidence supports the hypothesis that C sink limitations may be particularly important in cold, dry, and late-successional forests. For example, elevated concentrations of nonstructural C (e.g., starch and sugars) are frequently observed in colder environments or during drought (*8*, *12*). Additionally, free air CO_2 enrichment (FACE) experiments tend to show that increasing CO_2 concentration improves tree growth in early-stage forests but often not in mature forests, perhaps because of stronger nutrient limitations (*13*–*15*). But the relatively small scale and replication of FACE experiments, especially in mature forests, prevents general conclusions regarding the linkage between C source and sink dynamics in trees.

Colocated assessments of gross primary productivity (GPP; e.g., by eddy-covariance) and tree growth theoretically enable evaluation of the coupling between tree C assimilation and growth increment. Past studies adopting such an approach were nonetheless limited by dataset size (site number ≤5) and yielded contrasting findings, with no clear explanation of observed differences (*16*–*21*). The advent of large-scale, long-term networks of flux towers measuring C exchange across a diverse assemblage of biomes, in combination with a growing number of both on-site and global tree ring datasets, opens new opportunities to characterize C source-sink relationships at larger temporal and spatial scales. Here, we compiled a new dataset comprising eddy-covariance GPP records at 78 forest flux sites (table S1), together with on-site tree ring width chronologies at a subset of 31 sites ($RW_{on\text{-}site}$; table S2) as well as 1800 nearby regional ring width chronologies (RW_{region}). GPP and RW records were detrended to remove low-frequency signals (e.g., stand structure; tree age and size) and were aggregated such that records were representative of year-to-year variations of stand C assimilation and aboveground woody growth, respectively (*22*). This C assimilation and tree growth dataset extends across most of Europe and North America, encompassing a variety of forested biomes from semi-arid to boreal, and representing both angiosperm and gymnosperm tree species (Fig. 1, fig. S1, and table S3). We used this dataset to (i) quantify the strength of tree C source-sink relationships across

[1]School of Biological Sciences, University of Utah, Salt Lake City, UT, USA. [2]McMaster Centre for Climate Change, McMaster University, Hamilton, Ontario L8S 4K1, Canada. [3]School of Earth, Environment and Society, McMaster University, Hamilton, Ontario L8S 4K1, Canada. [4]School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA. [5]Laboratory of Tree-Ring Research, University of Arizona, Tucson, AZ, USA. [6]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA. [7]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, 91405 Orsay, France. [8]Institut Universitaire de France, 75231 Paris Cedex 05, France. [9]DISTAL, Alma Mater Studiorum, University of Bologna, Bologna, Italy. [10]Department of Geography, Indiana University, Bloomington, IN, USA. [11]USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR, USA. [12]Centre d'étude de la forêt, Université du Québec à Montréal, C.P. 8888, Succursale Centre-ville, Montréal, Quebec H3C 3P8, Canada. [13]Département Science et Technologie, Téluq, Université du Québec, Bureau 1105, Montréal, Quebec H2S 3L5, Canada. [14]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA. [15]Geology Institute, National Autonomous University of Mexico, Coyoacán, CDMX, Mexico. [16]Graduate School of Life and Environmental Sciences, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai 599-8531, Japan. [17]Department of Ecology, Montana State University, Bozeman, MT, USA. [18]Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. [19]College of Forest Resources and Environmental Science, Michigan Technological University, Houghton, MI, USA. [20]Department of Ecology, W. Szafer Institute of Botany, Polish Academy of Sciences, 31-512 Kraków, Poland.
***Corresponding author. Email: antoine.cabon@utah.edu**

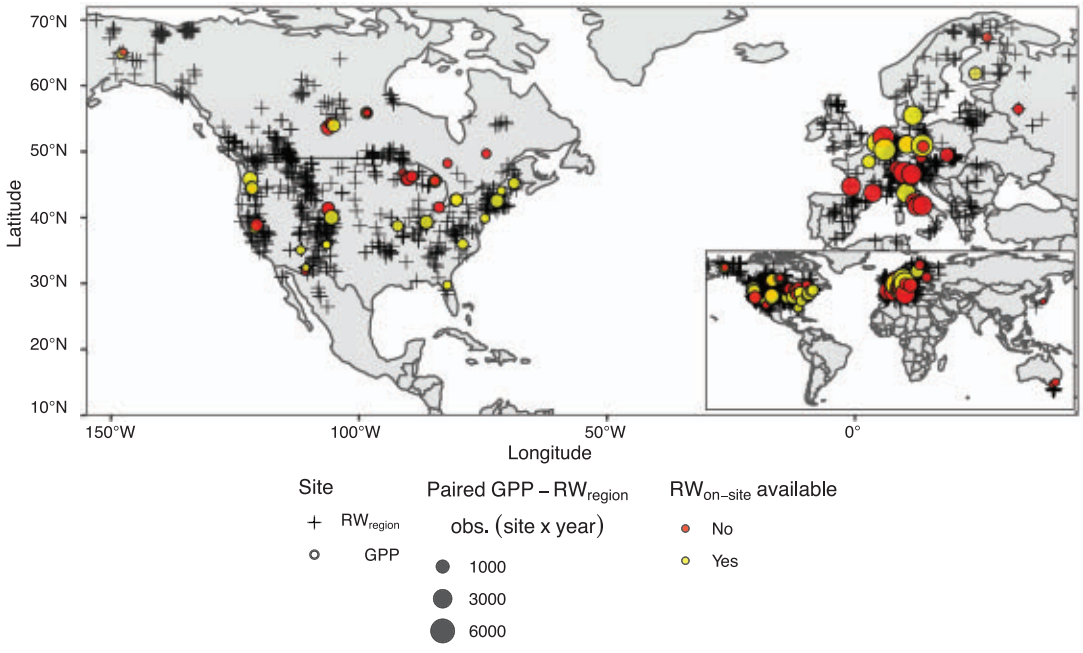


Fig. 1. Spatial distribution of gross primary production (GPP) and regional ring width (RW_{region}) sites used in this study. Crosses, RW_{region} sites; circles, GPP sites. Circle size denotes the number of RW_{region} site × year observations associated with each flux tower. GPP sites that further include on-site RW are shown in yellow; others are red.

biomes, (ii) identify the seasonality of these relationships, and (iii) explore their environmental drivers.

We first characterized C source and sink relationships at the regional scale by statistically accounting for the decrease of the correlation between GPP and RW_{region} (r_{region}) with increasing geographic and climatic distances, as well as with an index of species dissimilarity between sites (*22*) (fig. S2). As expected from reported tree growth synchrony over large distances (*23*), we observed sustained correlations up to ~500 km. We thus built on this widespread ecological feature to derive robust regional estimates of tree C assimilation and growth correlation, $r_{D=0}$, for theoretical colocated sites of identical climate and species composition (i.e., spatial distance, climatic distance, and species dissimilarity of 0), integrating over multiple time scales. We then complemented regional-scale analyses with paired GPP and on-site tree ring correlations ($r_{on\text{-}site}$; see annual GPP and RW series in fig. S3). The latter dataset has a lower sample size than the regional network but is model-free and therefore reduces the risk of methodological artifacts.

Both on-site and regional correlations showed an overall weak association between tree C assimilation and growth, with $r_{on\text{-}site}$ and $r_{D=0}$ reaching maxima of 0.26 and 0.38, respectively (Fig. 2, A and B). The observed difference between on-site and regional estimates could be offset by setting species dissimilarity to the average encountered for $RW_{on\text{-}site}$, resulting in a maximum regional correlation of 0.27 (*22*). RW_{region} observations

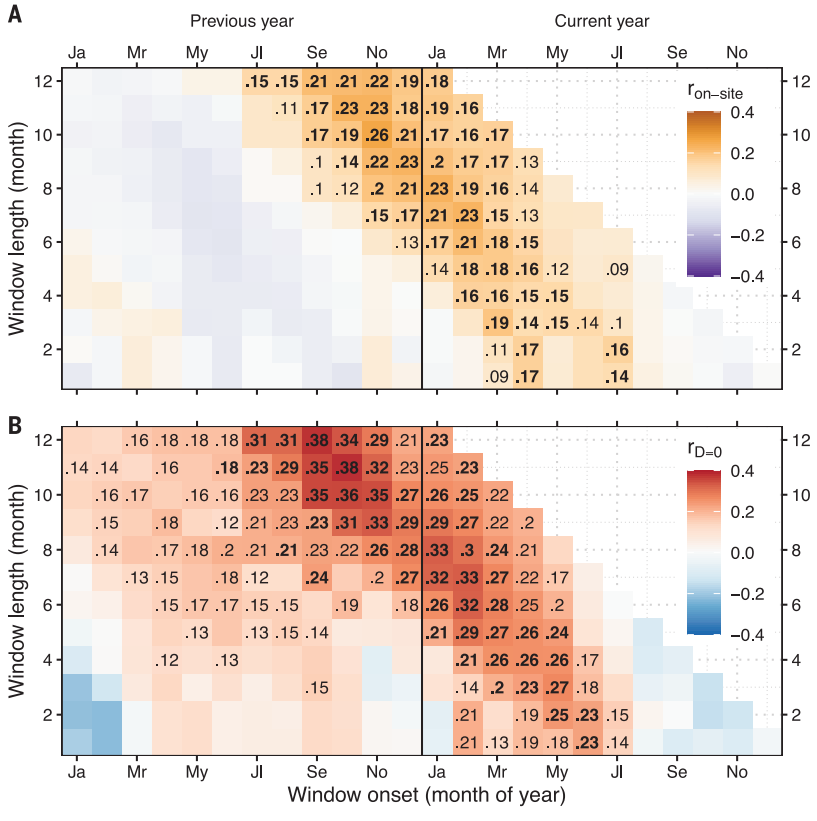


Fig. 2. Temporal structure of GPP versus RW correlations. (**A**) Seasonal on-site correlations ($r_{on\text{-}site}$). Each cell corresponds to the average correlation calculated between on-site RW and GPP summed over a time period defined by a window onset (from previous-year January to current-year December) and length (from 1 to 12 months). (**B**) Regional-based estimates of null distance correlations ($r_{D=0}$) modeled by eq. S1 (see fig. S2 for an illustration of the 12-month case from current-year January). Significant correlation values are displayed on top of corresponding cells (all p values < 0.05, except for boldface: p < 0.01).

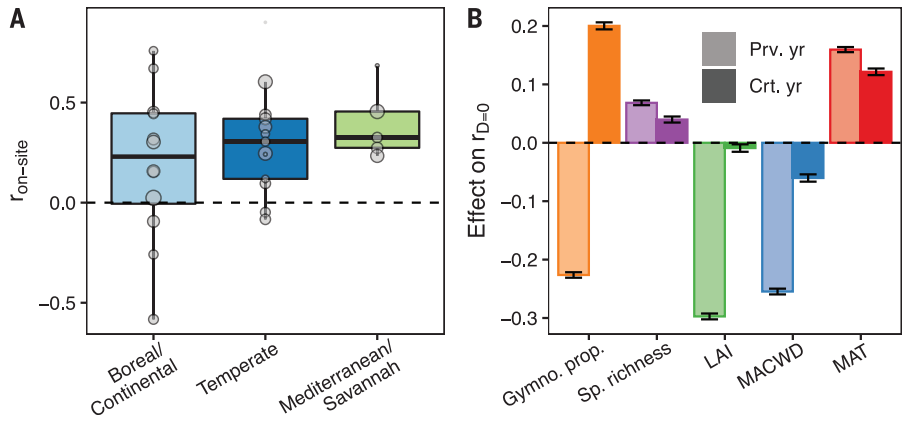


Fig. 3. Spatial variations and environmental drivers of GPP versus RW correlations. (**A**) Effect of biome on on-site correlations ($r_{on-site}$) observed in the period with highest correlation average (previous November through current October; nonsignificant). Boxes represent the median and first and third quartiles. Whiskers represent 1.5 times the interquartile range. Dots represent individual *r* values; dot size is proportional to the underlying number of observations. (**B**) Effect of stand structure and climatic variables on current- and previous-year regional-based estimates of null distance correlations ($r_{D=0}$). Error bars denote SE. All effects are highly significant [$p < 0.001$: gymnosperm proportion, species richness, mean annual climatic water deficit (MACWD), mean annual temperature (MAT)] except for that of leaf area index (LAI) on current-year correlations (nonsignificant).

partially build on the International Tree-Ring Data Bank, where sampling is often biased toward dominant and climate-sensitive trees (*24*). However, we find that this is unlikely to be an issue here, as dominant trees account for most of stand GPP and we statistically corrected for differences in climate (*22*). Overall, similar regional and on-site results show the suitability of regional RW data to quantify local GPP-RW correlations and broad agreement between the two approaches, both of which suggest a substantial decoupling between C assimilation and tree growth across multiple biomes.

On-site and regional GPP-RW correlations exhibited a similar temporal structure (*22*), with correlation magnitude increasing with the length of the GPP integration period and maximum correlations being found at the 10- and 12-month scales for $r_{on-site}$ and r_{region}, respectively (Fig. 2). This supports the often-implicit assumption that annual tree ring increments are most strongly related to annual carbon assimilation (*21*). Overall, RW was best correlated to GPP integrated over the period spanning previous-year September or November to current-year August, indicating a short temporal lag between C assimilation and tree growth, consistent with a previous study (*20*). This result suggests that despite estimated low C source limitation of tree growth overall, excess photosynthates are stored over winter (following radial growth cessation) and are allocated to the next year's growth. This phenomenon is often cited as a potential explanation for delayed climatic effects on tree growth and growth autocorrelation (*25*, *26*). Analysis of multiyear trends (table S4) nonetheless indicates weak association of RW and GPP at this scale, contrary to the hypothesis that C storage might lead to the convergence of tree growth and C assimilation over the long term (*27*).

We found large spatial variations in the strength of GPP-RW coupling (Fig. 3). Weighted deciles of maximum on-site *r* ranged from −0.08 to 0.60, consistent with previously reported values (*16*–*21*). These spatial variations imply a range of source versus sink limitations. We estimate that because of approximations and measurement errors, RW-GPP correlations between 0.7 and 0.9 would be expected under strong source control of tree growth (*22*). The high end of the observed correlation range ($0.6 \leq r_{on-site} \leq 0.9$: 10% of observations) thus appears reflective of substantial source limitation of tree growth at the corresponding sites, whereas the majority of sites display evidence consistent with sink limitations. Although we did not observe a biome effect on on-site correlations, regional-scale *r* was significantly related to several environmental factors (Fig. 3B). Specifically, gymnosperm proportion had a positive effect on current-year r_{region} but a negative one on previous-year r_{region}; this suggests that gymnosperm growth relies more directly on current-year and less on previous-year C assimilation than angiosperms, reflecting fundamental physiological differences between these two clades. A small but positive effect of species richness on r_{region} suggests a link between species diversity and C use efficiency (i.e., the ratio between net and gross primary production), which may arise as a result of increased complementarity with structural and functional heterogeneity (*28*). Decreasing r_{region} with increasing leaf area index indicates that closed-canopy forests, which under a given climate tend to be older and more nutrient-limited than open-canopy forests, are prone to stronger decoupling between C source and sink activity. This result agrees with observations that CO_2 growth fertilization tends to fade in older, nutrient-limited forests (*15*). Last, r_{region} was found to be positively related to site temperature and water availability, consistent with known biophysical controls of cambial activity and the ensuing prediction that sink limitations are stronger under colder and drier conditions (*6*–*9*). These combined results draw a clear picture that increasing resource limitation, aridity, and low temperatures promote C source-sink decoupling across a broad range of biomes.

Finally, decoupling of C assimilation and tree growth was further revealed by diverging climate sensitivities of these two processes (*22*) (Fig. 4). As anticipated from C assimilation and wood formation literature, GPP and RW both responded positively to temperature and water availability but were weakly correlated with photosynthetically active radiation (*29*, *30*). However, their seasonal variability differed markedly, indicating that fundamentally different physiological processes may limit C assimilation and tree growth. GPP responded mostly to spring and fall temperatures as well as to summer water availability, suggesting an important role of temperature-triggered leaf phenology controlling annual GPP (*31*). In contrast, RW appeared to be most strongly related to year-round water availability, with a weak positive temperature effect peaking in summer. This agrees with previous observations that tree growth is primarily and increasingly water-limited in the study regions (*29*) and is consistent with the central role of cell turgor in controlling cambial cell division and expansion (*7*, *11*). Overall, this analysis shows the large but contrasting climate sensitivity of the tree growth and photosynthesis proxies used here. This is contrary to the expectation that RW and GPP would have weaker but similar climate sensitivity if low RW-GPP were due primarily to large measurement errors. These results instead strongly suggest that weak control of C assimilation over tree growth is underpinned by fundamentally contrasting source and sink processes with diverging environmental sensitivities (*6*).

Taken together, our results provide consistent evidence for the pervasive influence of nonphotosynthetic processes on tree radial growth. This conclusion has major implications for projections of forest dynamics and feedbacks with the global C cycle and climate change, as most global vegetation models essentially simulate forest productivity and C sequestration as a linear function of C assimilation (*3*, *4*). Because, compared to C assimiliation, sink processes are relatively more

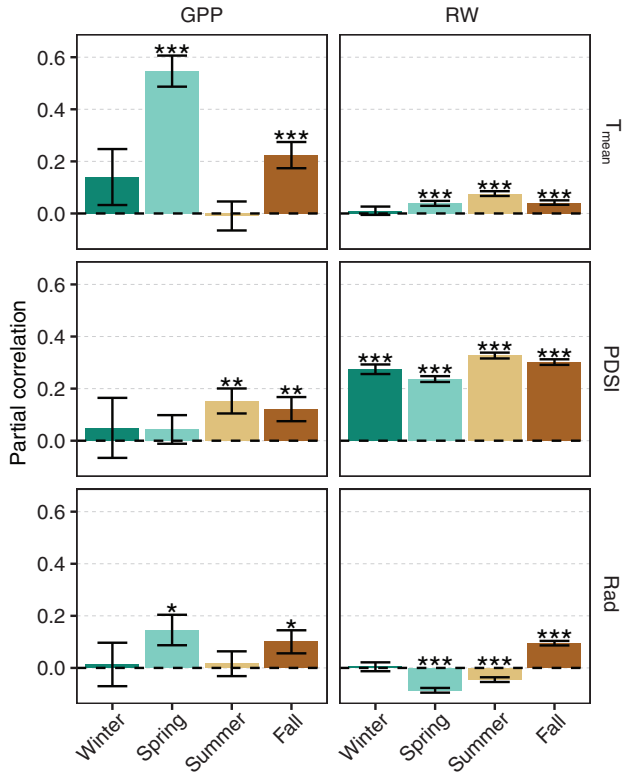


Fig. 4. GPP and RW_{region} climatic sensitivity. Climate-corrected partial correlations between GPP and RW_{region} and three climate variables [mean temperature (T_{mean}), Palmer's drought severity index (PDSI), and photosynthetically active radiation (Rad)] are shown at the 3-month scale over the period 1990–2015 (to the extent of series span). Error bars denote SE. *p < 0.05, **p < 0.01, ***p < 0.001.

sensitive to water availability and less to temperature constraints (Fig. 4), as well as not directly dependent on atmospheric CO_2 concentration, unaccounted-for and widespread sink limitations could lead to overestimating the positive effect of warming and CO_2 fertilization while underestimating the negative effect of increasing water stress on forest productivity. Overall, accounting for sink limitations on tree growth may lower projections of future forest C sequestration in many regions and could thus potentially compromise forests' potential for climate change mitigation. Hence, our results underscore that incorporation of sink-limited carbon allocation schemes in global vegetation models is urgently needed (*3*, *4*).

Our results nonetheless indicate a certain degree of interaction between C source and sink activities, as suggested by the weak but significantly positive correlations observed between GPP and RW, as well as their temporal and spatial variations. Such dynamic coupling between C assimilation and tree growth potentially reconciles contrasting observations of the prevalence of source versus sink limitations (*15*) and provides a bridge between current source-centered representations of tree growth and sink-driven schemes. Variations in the prevalence of source versus sink limitations to tree growth further highlight the importance of understanding their drivers (*5*). Our findings indicate that across biomes, the occurrence of sink limitations is highly consistent with known biophysical controls of cambial cell division, notably turgor-driven growth. Because turgor is a central mechanism of growth across scales and has a large potential for the integration of several relevant processes as well as parameter-parsimonious upscaling (*32*), the turgor-driven growth framework appears to be a promising way to progress toward developing mechanistic sink-limited schemes in vegetation models.

Key remaining uncertainties concern whether our results can be generalized to other biomes such as tropical forests, which are central to the global C cycle, and the dynamic nature of source and sink interactions. Likewise, characterizing the degree of C source and sink decoupling at decadal to centennial scales is relevant to climate change but currently remains elusive because of the temporal depth of C assimilation measurements. Source-sink decoupling over both short and longer time scales implies less C limitation of tree growth. Weak C limitation of tree growth under certain conditions nonetheless raises the question of the fate of excess C. Closing trees' C budget and elucidating drivers of C allocation to different sinks, specifically stem versus underground growth and C storage, thus emerges as a critical way forward (*14*).

REFERENCES AND NOTES

1. P. Friedlingstein *et al.*, *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).
2. T. A. M. Pugh *et al.*, *Biogeosciences* **17**, 3961–3989 (2020).
3. A. D. Friend *et al.*, *Ann. For. Sci.* **76**, 49 (2019).
4. S. Fatichi, S. Leuzinger, C. Körner, *New Phytol.* **201**, 1086–1095 (2014).
5. F. Babst *et al.*, *Trends Plant Sci.* **26**, 210–219 (2021).
6. C. Körner, *Curr. Opin. Plant Biol.* **25**, 107–114 (2015).
7. T. C. Hsiao, *Annu. Rev. Plant Physiol.* **24**, 519–570 (1973).
8. B. Muller *et al.*, *J. Exp. Bot.* **62**, 1715–1729 (2011).
9. R. L. Peters *et al.*, *New Phytol.* **229**, 213–229 (2021).
10. I. Cornut *et al.*, *For. Ecol. Manage.* **494**, 119275 (2021).
11. A. Cabon, R. L. Peters, P. Fonti, J. Martínez-Vilalta, M. De Cáceres, *New Phytol.* **226**, 1325–1340 (2020).
12. G. Hoch, C. Körner, *Glob. Ecol. Biogeogr.* **21**, 861–871 (2012).
13. T. Klein *et al.*, *J. Ecol.* **104**, 1720–1733 (2016).
14. M. Jiang *et al.*, *Nature* **580**, 227–231 (2020).
15. A. P. Walker *et al.*, *New Phytol.* **229**, 2413–2445 (2021).
16. A. V. Rocha, M. L. Goulden, A. L. Dunn, S. C. Wofsy, *Glob. Change Biol.* **12**, 1378–1389 (2006).
17. F. Babst *et al.*, *New Phytol.* **201**, 1289–1303 (2014).
18. N. Delpierre, D. Berveiller, E. Granda, E. Dufrêne, *New Phytol.* **210**, 459–470 (2016).
19. C. Pappas *et al.*, *Agric. For. Meteorol.* **290**, 108030 (2020).
20. A. Teets *et al.*, *Agric. For. Meteorol.* **249**, 479–487 (2018).
21. M. Lempereur *et al.*, *New Phytol.* **207**, 579–590 (2015).
22. See supplementary materials.
23. M. del Río *et al.*, *For. Ecol. Manage.* **479**, 118587 (2021).
24. S. Klesse *et al.*, *Nat. Commun.* **9**, 5336 (2018).
25. R. Zweifel, F. Sterck, *Front. For. Glob. Change* **1**, 9 (2018).
26. A. Gessler, K. Treydte, *New Phytol.* **209**, 1338–1340 (2016).
27. C. M. Gough, C. S. Vogel, H. P. Schmid, H. B. Su, P. S. Curtis, *Agric. For. Meteorol.* **148**, 158–170 (2008).
28. S. Mensah, R. Veldtman, A. E. Assogbadjo, R. Glèlè Kakaï, T. Seifert, *Ecol. Evol.* **6**, 7546–7557 (2016).
29. F. Babst *et al.*, *Sci. Adv.* **5**, eaat4313 (2019).
30. N. Delpierre *et al.*, *Agric. For. Meteorol.* **154–155**, 99–112 (2012).
31. J. Xia *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2788–2793 (2015).
32. A. Potkay, T. Hölttä, A. T. Trugman, Y. Fan, *Tree Physiol.* **42**, 229–252 (2022).

ACKNOWLEDGMENTS

We thank C. Hanson, S. Wharton, R. Brooks, and S. Klesse for contributing data to this study, as well as contributors at the International Tree-Ring Data Bank, FLUXNET, and AmeriFlux. **Funding:** Supported by USDA National Institute of Food and Agriculture, Agricultural and Food Research Initiative Competitive Program, Ecosystem Services and Agro-Ecosystem Management, grant 2018-67019-27850 (A.C., S.A.K., and W.R.L.A.); the David and Lucile Packard Foundation and NSF grants 1714972, 1802880, 2044937, and 2003017 (W.R.L.A.); NSF Ecosystem Science cluster grant 1753845, USDA Forest Service Forest Health Protection Evaluation Monitoring program grant 19-05, and DOE Environmental System Science program grant DE-SC0022052 (S.A.K.); Arctic Challenge for Sustainability II grant JPMXD1420318865 (M.U.); USDA National Institute of Food and Agriculture, Agricultural and Food Research Initiative Competitive Program grant 2017-67013-26191 (J.T.M.); and DOE Office of Biological and Environmental Research grant DE-SC0010611 and NSF Directorate for Biological Sciences grant 1241851 (D.J.M.). Funding for the AmeriFlux data portal was provided by the US Department of Energy Office of Science. **Author contributions:** Conceptualization: A.C., W.R.L.A. Methodology: A.C., W.R.L.A., S.A.K. Data contributions: All co-authors. Investigation: A.C., W.R.L.A., S.A.K. Visualization: A.C. Funding acquisition: W.R.L.A. Writing–original draft: A.C., W.R.L.A., S.A.K. Writing–review and editing: All co-authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All processed data used for the analyses are available on Dryad (DOI: 10.5061/dryad.15dv41nzt) and the code is available on Zenodo (DOI: 10.5281/zenodo.6033963).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm4875
Materials and Methods
Figs. S1 to S5
Tables S1 to S4
References (*33*–*55*)

Submitted 21 September 2021; accepted 16 February 2022
10.1126/science.abm4875

SOLAR CELLS

Scalable processing for realizing 21.7%-efficient all-perovskite tandem solar modules

Ke Xiao[1,2], Yen-Hung Lin[3], Mei Zhang[1], Robert D. J. Oliver[3], Xi Wang[4,5], Zhou Liu[1], Xin Luo[1,2], Jia Li[5], Donny Lai[5], Haowen Luo[1], Renxing Lin[1], Jun Xu[2], Yi Hou[4,5], Henry J. Snaith[3]*, Hairen Tan[1]*

Challenges in fabricating all-perovskite tandem solar cells as modules rather than as single-junction configurations include growing high-quality wide-bandgap perovskites and mitigating irreversible degradation caused by halide and metal interdiffusion at the interconnecting contacts. We demonstrate efficient all-perovskite tandem solar modules using scalable fabrication techniques. By systematically tuning the cesium ratio of a methylammonium-free 1.8–electron volt mixed-halide perovskite, we improve the homogeneity of crystallization for blade-coated films over large areas. An electrically conductive conformal "diffusion barrier" is introduced between interconnecting subcells to improve the power conversion efficiency (PCE) and stability of all-perovskite tandem solar modules. Our tandem modules achieve a certified PCE of 21.7% with an aperture area of 20 square centimeters and retain 75% of their initial efficiency after 500 hours of continuous operation under simulated 1-sun illumination.

Monolithic all-perovskite tandem solar cells show great promise for large-scale photovoltaic (PV) applications with the advantage of low-cost solution processing (*1*–*3*). However, certified power conversion efficiencies (PCEs), which can reach up to 26.4% (*4*, *5*), have only been achieved in small-area devices with lab-scale spin-coating techniques that limit scalability. To enable large-area fabrication of perovskite films, deposition techniques such as spray coating (*6*), inkjet printing (*7*), blade coating (*8*, *9*), slot-die coating (*10*, *11*), and vacuum evaporation (*12*) have been reported. Solution-based fabrication routes all involve solvent engineering to modulate the crystallization dynamics, but the solvent systems that are in use now for scalable coating of state-of-the-art ~1.5-eV-bandgap perovskite films are incompatible with those of the ~1.8-eV-bandgap wide-bandgap (WBG) perovskites that are needed for all-perovskite tandem solar modules (*10*). The higher bromide concentration in WBG perovskites leads to variations in crystallization kinetics, and precursor solutions are limited by the low solubility of lead and cesium bromide salts (*13*). These constraints hinder the scalable fabrication of high-quality WBG perovskites for all-perovskite tandem solar modules (*14*, *15*).

[1]National Laboratory of Solid State Microstructures, Jiangsu Key Laboratory of Artificial Functional Materials, College of Engineering and Applied Sciences, Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing 210023, China. [2]School of Electronics Science and Engineering, Nanjing University, Nanjing 210093, China. [3]Clarendon Laboratory, Department of Physics, University of Oxford, Oxford OX1 3PU, UK. [4]Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585, Singapore. [5]Solar Energy Research Institute of Singapore (SERIS), National University of Singapore, 7 Engineering Drive 1, Singapore 117574, Singapore.
*Corresponding author. Email: hairentan@nju.edu.cn (H.T.); henry.snaith@physics.ox.ac.uk (H.J.S)

Another challenge in fabricating perovskite solar modules is linked to the reaction of halides and metal electrodes at the interconnecting subcells. The interdiffusion between the perovskite absorber and metal creates deep defect states at either the interface or the bulk of the perovskites (*16*–*19*). Unlike small-area perovskite solar cells (PSCs), perovskite solar modules require a three-step laser or mechanical scribing (namely P1, P2, and P3) to connect the subcells in series (*20*–*25*). The direct contact between perovskites and metal electrodes at the interconnecting areas between cells leads to subsequent halide-metal interdiffusion and limits the performance and stability of the modules (*26*). If the tandem-cell recombination junction is a highly conductive transparent conducting oxide (*27*–*29*), which is often the case, then metal-to–recombination layer contact will also lead to short-circuiting of one or both of the subjunctions. Injecting two-dimensional (2D) barrier materials, such as 2D nanostructured graphitic carbon nitride, between subcells can inhibit such interdiffusion (*26*). However, the poor electronic property of these 2D materials adversely causes an undesirable, large hysteresis in the solar modules. Furthermore, adding a 2D diffusion barrier layer between subcells complicates the overall fabrication process and reduces the geometric fill factor (GFF) because it requires two more independent processes (spin-coating and injecting) and necessitates a much wider space gap between subcells. A one-step process to deposit a thin conformal diffusion barrier (CDB) would not only improve cost effectiveness but also reduce the cell-to-module efficiency gap.

In this work, we controlled the homogeneity of crystallization in WBG perovskites over large areas by tuning the content of monovalent inorganic cation cesium. This strategy enabled the fabrication of 1-cm^2 all-perovskite tandem solar cells with a steady-state PCE of 24.8% through scalable processing techniques. A CDB consisting of atomic-layer-deposited SnO_2 (ALD-SnO_2) served as both the vertical electron extractor and the lateral diffusion barrier between interconnecting subcells. The CDB inhibited halide-metal interdiffusion and avoided the reaction between perovskites and metal electrode. Using the ALD-SnO_2–based CDB, we demonstrated all-perovskite tandem solar modules with a certified PCE of 21.7% (aperture area of 20.25 cm^2). Encapsulated tandem solar modules retained 75% of their initial performance after aging for 500 hours at the maximum power point (MPP) operation under simulated 1-sun illumination in an ambient condition.

We first attempted to blade coat the WBG perovskite films with a composition of $Cs_{0.2}FA_{0.8}PbI_{1.8}Br_{1.2}$ (where FA is formamidinium), which was used for the spin-coating process in our previous works (*2*, *3*). The relatively volatile solvent 2-methoxyethanol (2-ME) favors rapid deposition of a uniform perovskite film with low bromide content (*21*, *30*). However, undesirable crystal precipitation occurred when adding WBG perovskite precursors into a 2-ME and dimethyl sulfoxide (DMSO) mixed solvent, given the limited solubility of lead and cesium bromide salts in 2-ME (fig. S1). A stable and transparent precursor solution was obtained when the WBG perovskite precursors were added to the coordinating *N*,*N*-dimethylformamide (DMF) and DMSO mixed solvent (fig. S1). We used several solvent-quenching methods—such as vacuum flashing, hot casting, and gas quenching—to remove excess DMF and DMSO solvent after the blade-coating step, but none of these methods resulted in dense and uniform perovskite layers (fig. S2). We found that optimizing a range of blading parameter spaces—that is, blade speed, quenching gas pressure, and the gap between the blade and the substrate—was not enough to obtain high-quality, uniform $Cs_{0.2}FA_{0.8}PbI_{1.8}Br_{1.2}$ perovskite films (fig. S3). We also attempted to blade-coat films with neat FA cation and a composition of $FAPbI_{1.8}Br_{1.2}$; however, the resulting films exhibited an obviously nonperovskite δ phase (fig. S4).

We found that the crystal nucleation rate could be controlled by finely tuning the Cs content (denoted as x in the $Cs_xFA_{1-x}PbI_{1.8}Br_{1.2}$ formula) in conjunction with a gas-assisted blade-coating technique (Fig. 1A). During the gas-quenching step, the nucleation process was initiated under supersaturation conditions, which was evident from the observation that the wet perovskite film turned brown with increased Cs content (fig. S5). We observed the enlargement of grain sizes and the flattening of the film surface with higher Cs incorporation (up to 35 mol %) into the perovskite. For film where x is equal to 0.35 ($Cs_{0.35}FA_{0.65}PbI_{1.8}Br_{1.2}$),

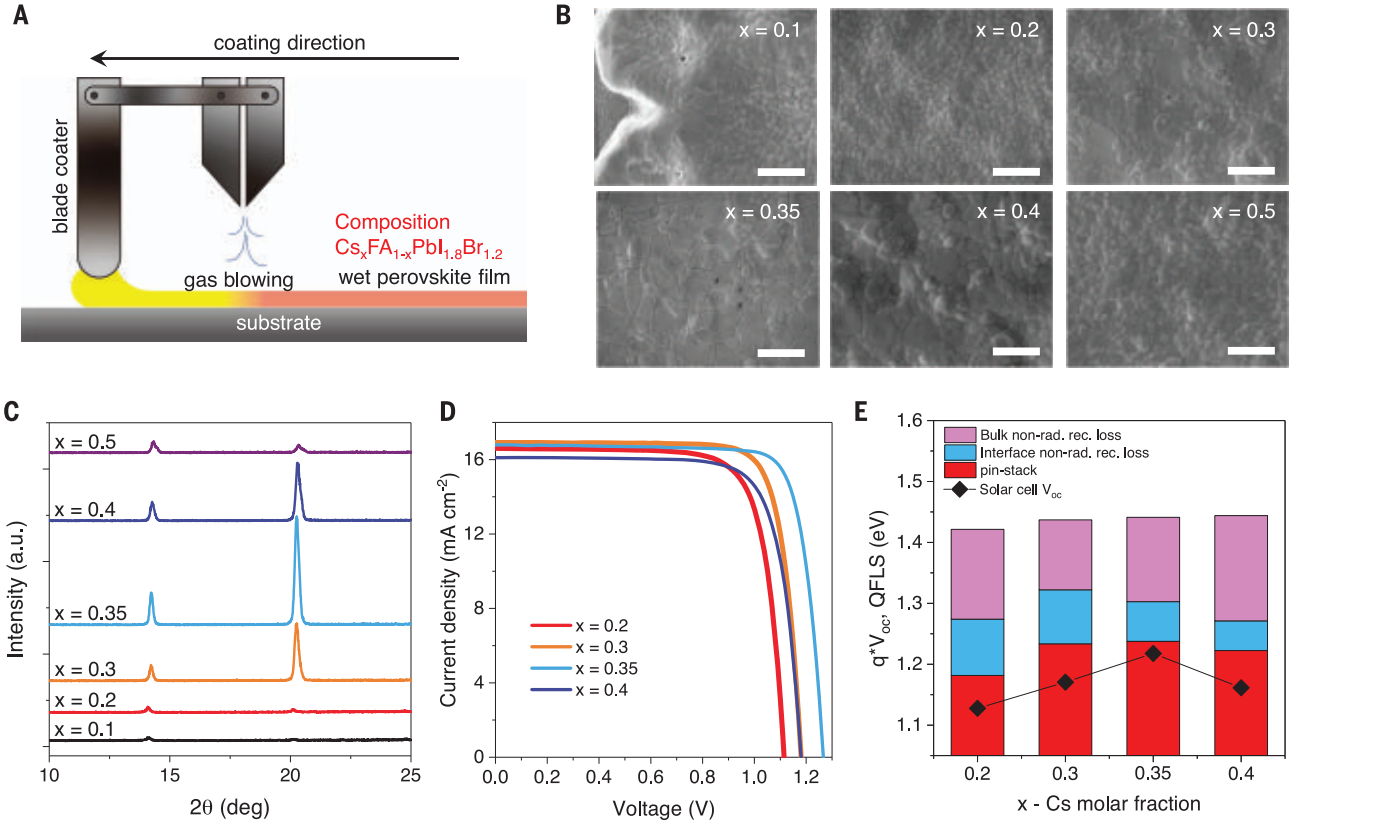


Fig. 1. Fabrication of $Cs_xFA_{1-x}PbI_{1.8}Br_{1.2}$ WBG perovskite films using blade coating. (**A**) Schematic illustration of gas-assisted blade coating. (**B** and **C**) SEM images (B) and x-ray diffraction patterns (C) of $Cs_xFA_{1-x}PbI_{1.8}Br_{1.2}$ perovskite films. The scale bars in the SEM images are 1 μm. a.u., arbitrary units. (**D**) *J-V* curves of champion $Cs_xFA_{1-x}PbI_{1.8}Br_{1.2}$ PSCs. (**E**) QFLS calculated from the PLQY for the respective perovskite films and the perovskites with transport layer stacks that were investigated in the study. The Shockley-Queisser radiative limit and the experimental V_{oc} are plotted for each composition. non-rad. rec. loss, nonradiative recombination loss.

we observed the largest grain sizes with a uniform surface, as observed from scanning electron microscopy (SEM) and atomic force microscopy (AFM) images (Fig. 1B and fig. S6). It should be noted that the grains or grain boundaries seen in SEM and AFM observations are only features of morphological domains or domain boundaries (*31*). The $Cs_{0.35}FA_{0.65}PbI_{1.8}Br_{1.2}$ perovskite film also exhibited the highest crystallinity, as indicated by the x-ray diffraction patterns (Fig. 1C and fig. S7A).

The grazing-incidence wide-angle x-ray scattering and the corresponding diffraction mottling intensity profiles integrated along the ring with scattering vector (q) equal to 1.0 Å^{-1} are shown in fig. S8. For blade-coated $Cs_{0.35}FA_{0.65}PbI_{1.8}Br_{1.2}$ perovskite films, the diffraction mottling at the azimuth angles of 45° (135°) and 90° became distinct compared with the ambiguous diffraction mottling at 60° (120°) for spin-coated $Cs_{0.2}FA_{0.8}PbI_{1.8}Br_{1.2}$ perovskite films. This new pattern implied that a distinct stacking orientation developed along with (100) crystallographic planes for the WBG perovskite films deposited through gas-assisted blade coating. Large crystal grains grown perpendicular to the substrate throughout the whole film were observed in blade-coated perovskite films (fig. S8). By contrast, smaller grains with multiple grain boundaries throughout the film were formed for spin-coated films. We observed that the grain orientation and crystallinity of blade-coated films were strongly related to the Cs content, which had only a weak impact on spin-coated films (fig. S9).

We noted a linear increase in the bandgap of $Cs_xFA_{1-x}PbI_{1.8}Br_{1.2}$ perovskites when x was increased up to 0.35. However, nonlinearity in the optical bandgap was seen when x was greater than 0.35 (fig. S10), which may have been caused by phase segregation in the perovskite films (fig. S7B). Perovskites with x less than 0.3 suffered from light-induced phase segregation under high illumination intensities (i.e., 10 suns), which was evident from the multiple emission peaks observed in the steady-state photoluminescence (PL) spectra (fig. S11). The reduced photoinduced phase segregation in film with x equal to 0.35 could be due to the reduced lattice strain with a higher Cs ratio (*32*) and/or the change in thermodynamics related to the intermixing of ions or ion migration within the crystal lattice (*33*, *34*).

We then evaluated the effect of Cs content on the optoelectronic properties and PV performance of inverted positive-intrinsic-negative (p-i-n)–structured WBG PSCs. The $Cs_{0.35}FA_{0.65}PbI_{1.8}Br_{1.2}$ perovskite films showed the longest carrier lifetime, as determined by PL decay (fig. S12). The performance distribution, external quantum efficiency (EQE), and steady-state power output of various $Cs_xFA_{1-x}PbI_{1.8}Br_{1.2}$–based devices are summarized in fig. S13, suggesting an optimal composition of $Cs_{0.35}FA_{0.65}PbI_{1.8}Br_{1.2}$. The champion $Cs_{0.35}FA_{0.65}PbI_{1.8}Br_{1.2}$ device achieved a PCE of 17.2%, with an open-circuit voltage (V_{oc}) of 1.266 V, a short-circuit current density (J_{sc}) of 16.8 mA cm^{-2}, and a fill factor (FF) of 80.9% (Fig. 1D and table S1). The blade-coated devices (x = 0.35) exhibited comparable performance to the spin-coated ones (x = 0.20; fig. S14), and the spin-coated devices with various Cs contents showed similar performance (fig. S15).

To understand how the Cs content (x = 0.2 to 0.4) affects the performance of blade-coated cells, we measured the transient photovoltage decay under open-circuit conditions (fig. S16). The photovoltage-decay lifetime τ of the $Cs_{0.35}FA_{0.65}PbI_{1.8}Br_{1.2}$ device (86 μs) was higher than that of the $Cs_{0.2}FA_{0.8}PbI_{1.8}Br_{1.2}$ device (31 μs), indicating a reduced recombination when the ratio of Cs was increased from 0.2 to

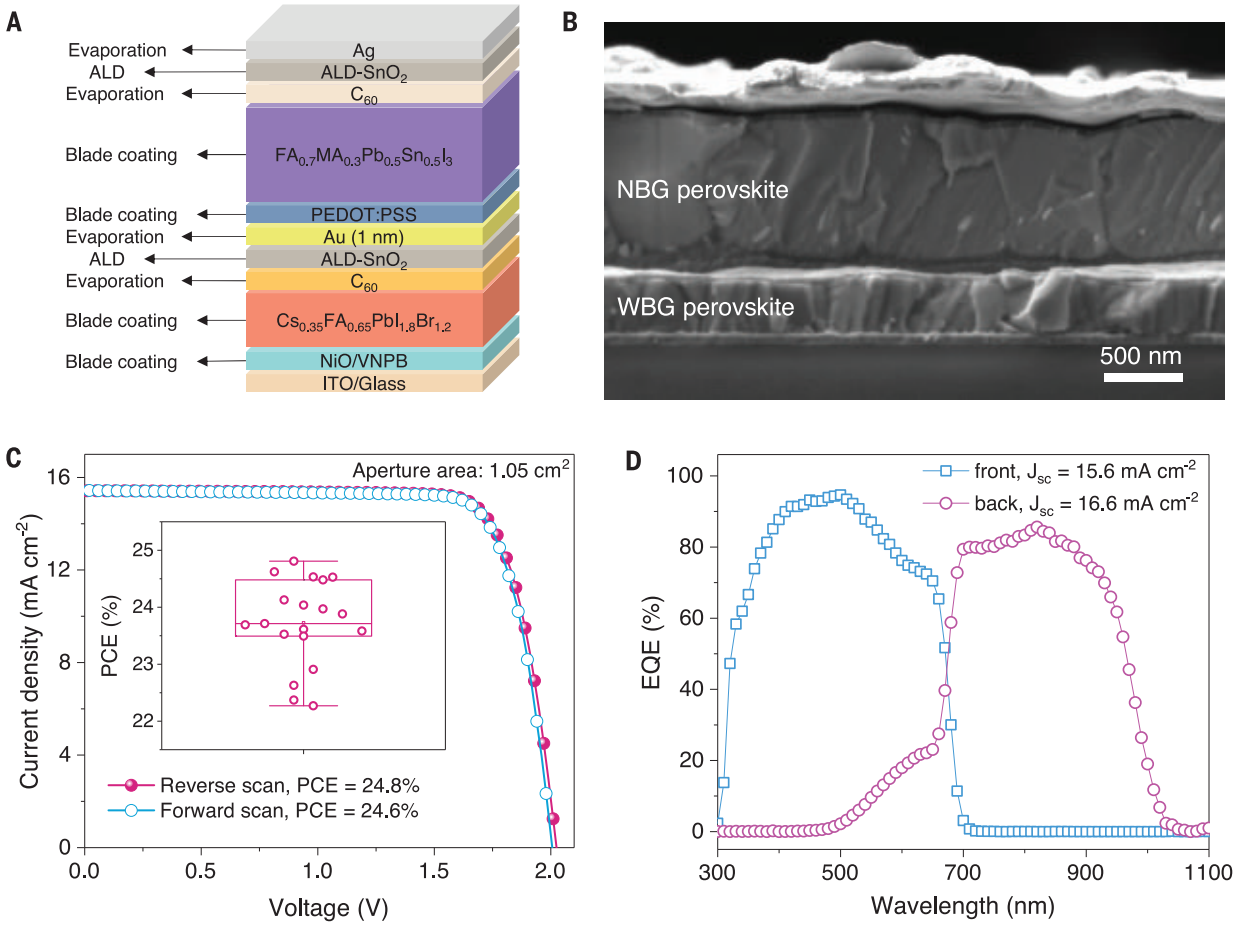


Fig. 2. Full fabrication of all-perovskite tandem solar cells using scalable techniques. (**A**) Configuration schematic of an all-perovskite tandem device fully fabricated using scalable techniques. The processing technique for each layer is indicated to the left. (**B**) Cross-sectional SEM image of the all-perovskite tandem device. (**C**) *J-V* curves of the champion tandem solar cell (aperture area of 1.05 cm^2) deposited on a 2.5-cm–by–2.5-cm substrate. The inset shows the PCE distribution of 19 devices, where the boxes and whiskers represent the standard deviation and the maximum and minimum of the distributions, respectively. (**D**) EQE curves of the champion device. The front and back subcells have integrated J_{sc} values of 15.6 and 16.6 mA cm^{-2}, respectively.

0.35. To further elucidate V_{oc} improvement of the PSCs with various Cs ratios, we investigated the PL quantum yield (PLQY) of isolated perovskite thin films and device stacks that had both charge-transporting layers present (fig. S17). These measured PLQY values were then used to derive the quasi-Fermi level splitting (QFLS) in the respective perovskite materials and p-i-n stacks (Fig. 1E). Notably, the calculated QFLS values of the p-i-n stacks corroborated well with the V_{oc} of the WBG PSCs and suggested that nonradiative recombination losses were reduced by increasing Cs to an x of 0.3 to 0.35.

We then investigated the uniformity of $Cs_{0.35}FA_{0.65}PbI_{1.8}Br_{1.2}$ perovskite films that were blade coated over a large area (6-cm–by–6-cm substrate; fig. S18). We patterned eight solar cells over a single substrate with each device pixel having an aperture area of 2.2 cm by 1.125 cm (2.475 cm^2), as illustrated in the inset of fig. S19A. All eight devices showed nearly identical device performance with a minor PCE standard deviation of 0.03% (fig. S19 and table S2). The narrowly distributed device performance indicates the effectiveness of the blade-coating method to achieve large-area uniformity for WBG perovskite films. By contrast, the devices distributed on the 6-cm–by–6-cm substrate by spin coating exhibited larger variations in performance among eight pixels (PCE standard deviations of 0.36 and 0.63% for $Cs_{0.35}FA_{0.65}PbI_{1.8}Br_{1.2}$ and $Cs_{0.2}FA_{0.8}PbI_{1.8}Br_{1.2}$, respectively). It should be noted that tuning the Cs content is likely not the only way to achieve high-quality, uniform WBG perovskite films, so further work in optimizing the blading parameter space, precursor solvent, and additives may also lead to similar or even better-quality films.

To realize tandem solar cells with scalable manufacturing techniques, we used blade coating to replace spin coating in all of the solution-based processes, including the fabrication of narrow bandgap (NBG) perovskite film ($MA_{0.3}FA_{0.7}Pb_{0.5}Sn_{0.5}I_3$, where MA is methylammonium) and hole transport layers (Fig. 2A). All other material layers were deposited through either thermal evaporation or ALD, which are both scalable processes already used in PV manufacturing. The blade-coated NBG PSCs (0.049 cm^2) delivered a champion PCE of 19.0% (steady-state PCE of 19.0%) and good reproducibility (fig. S20). Monolithic all-perovskite tandem solar cells, consisting of a ~400-nm-thick WBG perovskite and a ~950-nm-thick NBG perovskite (Fig. 2B), were fabricated entirely with scalable techniques. The current density–voltage (*J-V*) and EQE curves of the best-performing tandem device with an aperture area of 1.05 cm^2 are presented in Fig. 2, C and D, respectively. An average PCE of 23.7 ± 0.7% was obtained from 19 devices (inset of Fig. 2C). The champion device had a PCE of 24.8% from the reverse scan, with a V_{oc} of 2.025 V, a J_{sc} of 15.4 mA cm^{-2}, and a FF of 79.4%. A highest efficiency of 25.1% (average PCE of 24.2 ± 0.6% from 44 devices) was obtained for 0.049-cm^2 tandem devices (fig. S21). The relatively modest PCE difference between

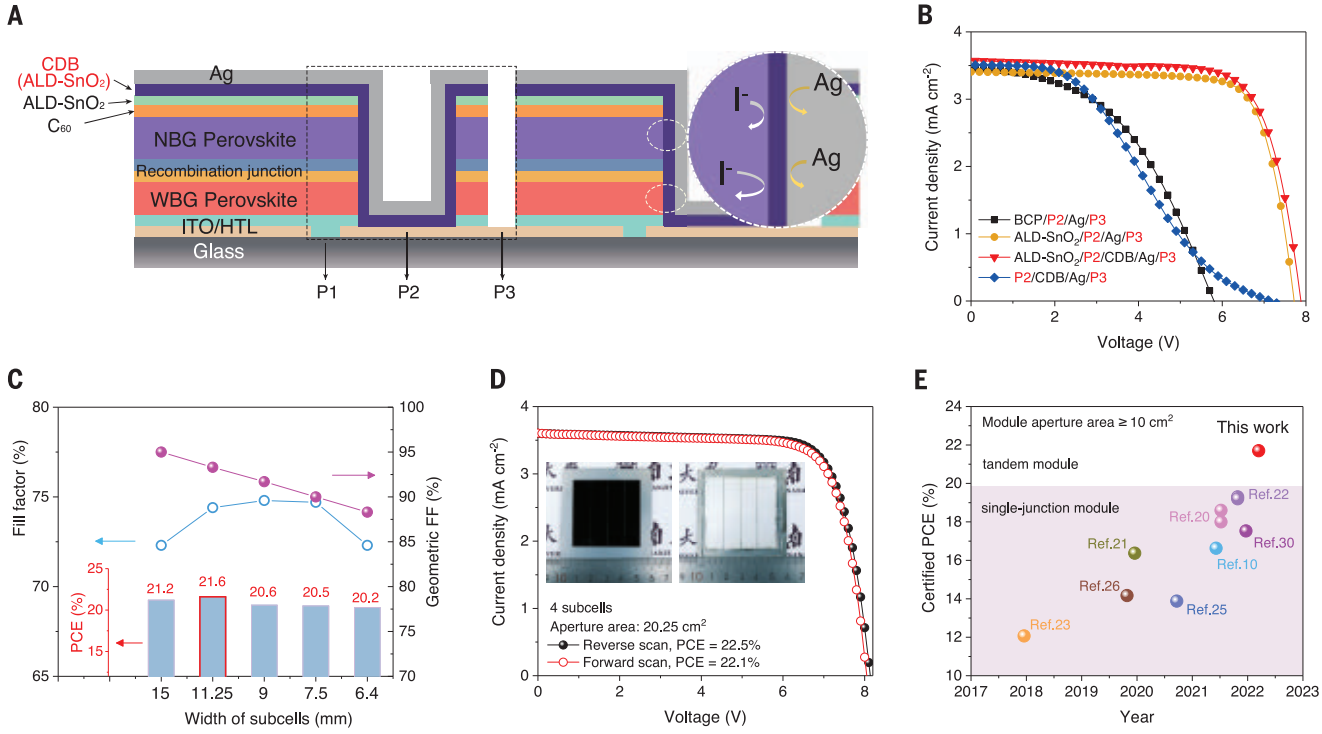


Fig. 3. All-perovskite tandem modules. (**A**) Schematic diagram of the structure of the series-connected all-perovskite tandem module with CDB to prevent ion diffusion. HTL, hole transport layer. (**B**) *J-V* curves of different module configurations. BCP, bathocuproine. (**C**) Relationship between the FF, GFF, and efficiency of modules. The arrows indicate GFF (purple), FF (blue), and PCE (red). (**D**) *J-V* curves of the champion all-perovskite tandem module (aperture area of 20.25 cm^2, four subcells in series). The inset shows photos of the front side (left) and back side (right) of the module. (**E**) Summary of publicly reported, independently certified PCEs of perovskite solar modules (aperture area ≥10 cm^2).

the 1.05- and 0.049-cm^2 cells suggests good scalability for all-perovskite tandem solar cells.

We fabricated all-perovskite tandem solar modules on 6-cm–by–6-cm substrates. The long-term stability and efficiency degradation of perovskite tandem modules is attributed in part to the interfacial halide-metal electrode reaction at the P2-scribed regions between the interconnecting subcells (Fig. 3A). To address this challenge, we devised an electrically conductive CDB by depositing ~10-nm-thick ALD-SnO_2 after P2 scribing to avoid interdiffusion and reaction between the perovskite and metal electrode (figs. S22 and S23). The CDB layer not only reduces the module manufacturing complexity but also enables a much narrower P2-scribed region (thus higher GFF and module efficiency) compared with injecting a wide insulating 2D barrier material at the interconnecting regions (*26*). We noted that another ALD-SnO_2 protective layer (~10 nm) deposited on C_{60} before the P2 scribing was essential to enable the P2 process to be performed under ambient conditions (Fig. 3B). We speculate that this compact protective layer prevented NBG perovskite from oxidation during exposure to ambient air (*2*).

In comparison, all of the modules that did not have the ALD-SnO_2 protective layer before performing the P2 process showed inferior performance. After the P2 process, the ALD-SnO_2–based CDB layer prevented direct contact between the metal electrode and the perovskite absorber at the P2-scribed regions. It also prevented direct contact between the metal electrode and the conductive poly(3,4-ethylenedioxythiophene) polystyrene sulfonate (PEDOT:PSS) in the recombination layer, which could otherwise lower the shunt resistance. The current-voltage measurements, performed at the ITO/ALD-SnO_2/Ag junction (where ITO is indium tin oxide), showed good ohmic contact with low vertical resistance (fig. S24), suggesting that this semiconducting CDB layer allowed effective electrical interconnection between subcells. The CDB layer improved both the efficiency and reproducibility of the all-perovskite tandem solar modules (fig. S25A).

We investigated how the subcell width would affect the performance of tandem modules. Increasing the subcell width allows for a higher GFF and thus potentially higher module efficiency, but this strategy adversely introduces a higher series resistance and hence reduces the FF. We fabricated modules that had three to seven subcells with a fixed total area, corresponding to subcell widths ranging from 15 to 6.4 mm (fig. S25B). The GFF is calculated to decrease linearly from 95.0 to 88.3% as the number of subcells increases (Fig. 3C and fig. S26). The optimal performance was achieved for the four-subcell tandem modules with a width of 11.25 mm and a GFF of 93.3% (Fig. 3C and table S3). We expect that laser scribing (*35*, *36*), instead of mechanical scribing, for P2 and P3 processings could allow for an even higher GFF and thus higher module efficiency, given that the subcell widths of tandems that are used herein are much larger than those reported in single-junction perovskite solar modules (table S4).

We fabricated 50 all-perovskite tandem solar modules with CDB and with a subcell width of 11.25 mm, which showed an average PCE of 20.9% (fig. S25A). The performance of single-junction WBG and NBG modules is summarized in fig. S27 and table S5. The champion tandem module exhibited a high PCE of 22.5% under reverse scan, with a V_{oc} of 8.137 V, a J_{sc} of 3.60 mA cm^{-2}, and a FF of 76.8% (Fig. 3D). Considering a GFF of 93.3%, the active-area efficiency of the tandem module reached 24.1% (fig. S28). The tandem module showed a minor hysteresis between the reverse and forward scans (22.5 versus 22.1%) and a steady-state PCE of 22.5% at a MPP voltage of 6.7 V measured over 3 min (fig. S29).

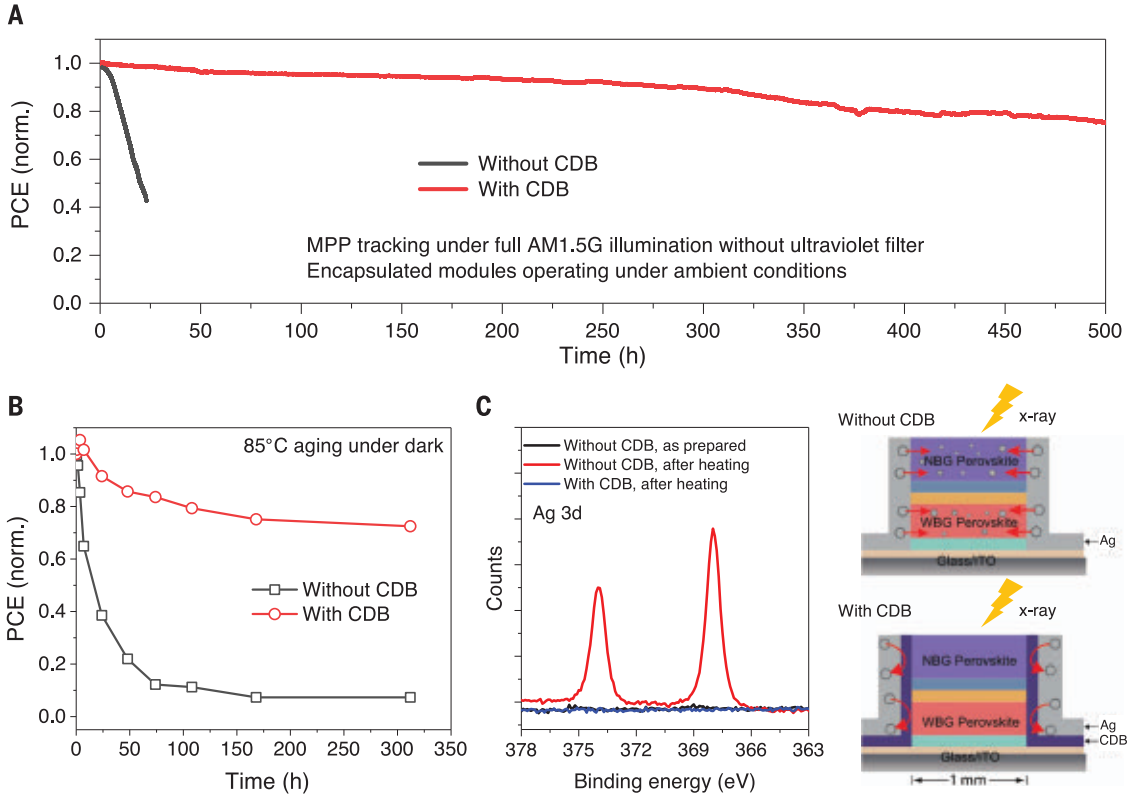


Fig. 4. Stability of all-perovskite tandem solar modules. (**A**) Continuous MPP tracking of an encapsulated tandem module over 500 hours under full simulated AM1.5 G illumination [100 mW cm^{-2}, light-emitting diode (LED) simulator] in ambient air with a humidity of 30 to 50%. norm., normalized. (**B**) Thermal stability tracking of the tandem modules heated on an 85°C hotplate in a N_2 glove box. (**C**) The Ag 3d XPS spectra of the perovskite surfaces after peeling off the C_{60}/ALD-SnO_2/Ag stacks. The x-ray beam is located at the center of the exposed perovskite surface after peeling off the C_{60}/ALD-SnO_2/Ag stacks.

A single tandem module under illumination could steadily power a cooling fan (fig. S29 and movie S1), and six tandem modules connected in parallel could charge a smartphone (movie S2). One module was sent to an accredited independent PV calibration and measurement laboratory (Japan Electrical Safety and Environment Technology Laboratories) for certification. The module delivered a certified PCE of 21.7% (fig. S30), which has been included in recent solar cell efficiency tables (version 59) (*5*). The certified 21.7% PCE of the tandem module surpasses those of single-junction perovskite solar modules with areas >10 cm^2 (Fig. 3E and table S4). The PCE of our tandem cell (~1 cm^2) using scalable fabrication delivered a performance comparable to that of tandem cells made by spin coating (table S6).

The encapsulated modules with CDB maintained their initial PCE after dark storage for 1778 hours under ambient conditions with a relative humidity of ~40% (fig. S31). We also tested the operating stability of encapsulated modules in ambient conditions under constant simulated 1-sun AM1.5 global (G) solar spectrum illumination. The module with CDB maintained 75% of its initial PCE (22.1%; fig. S32) after 500 hours of MPP tracking, whereas the module without CDB degraded to less than 50% of its initial PCE after 20 hours (Fig. 4A). We reasoned that the performance degradation of the module with CDB was not mainly attributed to the WBG perovskite; the single-junction WBG module could maintain 95% of its initial PCE after 450 hours of MPP tracking (fig. S33). Two reasons for why the degradation of the tandem modules with CDB is faster than that of the single-junction WBG module could be (i) Au as a recombination layer could have diffused into the NBG perovskite layer, leading to the formation of defect states at the perovskite interface or in the bulk (*18*), and (ii) the reaction at the PEDOT:PSS/NBG perovskite interface may have led to poorer charge extraction (*37*).

We tracked the thermal stability of encapsulated modules by heating at 85°C in a N_2 glove box. The module without CDB degraded down to 10% of the initial PCE after 312 hours, whereas the module with CDB still maintained >70% of its initial performance (Fig. 4B). For modules without CDB, erosion of the metal electrode was observed after thermal aging at the interconnecting regions (fig. S34). We speculated that the erosion was induced by the halide-metal interdiffusion at the interconnecting regions (lateral edges of the subcells in the module) because of the direct contact between perovskite and metal. The halide-metal interdiffusion would have two negative effects: (i) Metal diffusion into the perovskite absorber could degrade the perovskite and increase charge carrier recombination, and (ii) the halide diffusion into the electrode could corrode the metal and reduce its electrical conduction.

To investigate the halide-metal diffusion, we reduced the subcell width in a tandem module to obtain 1-mm-wide grid cells at the P2 processing step to monitor the I-Ag interdiffusion through x-ray photoelectron spectroscopy (XPS) characterization (fig. S35). The devices were then subjected to heating at 85°C in a N_2 glovebox for 24 hours. For modules without CDB, Ag was detected by XPS in perovskite films when the x-ray beam was placed on the perovskite film surface after peeling off the multiple top layers of C_{60}/ALD-SnO_2/Ag (Fig. 4C and fig. S35). This result indicated that Ag diffused laterally from the P2 region into the perovskite absorbers, whereas the halides (I^-) diffused into the edge metal electrode where they had direct contact with the perovskite (fig. S36). By contrast, for the CBD layer containing modules, no obvious signals of Ag (or I) were detected in the perovskite layer (or edge metal electrode), indicating that the I-Ag interdiffusion was effectively suppressed (*18*).

To intuitively elucidate the lateral interdiffusion between the perovskite and metal electrode, we further fabricated the following structures: (i) glass/ITO/~400-nm-thick WBG perovskite/Ag and (ii) glass/ITO/WBG perovskite/CDB/Ag, similar to the lateral structure in P2-scribed regions. After thermal aging at 85°C, we performed elemental analysis using SEM with energy-dispersive x-ray (fig. S37). The I-Ag interdiffusion occurred across the direct perovskite/electrode contact and extended over

the entire perovskite/metal layers, whereas the halide-metal interdiffusion was largely hindered by the use of CDB. Br-Ag interdiffusion was not observed (figs. S36 and S37), possibly because of the larger electronegativity of Br and thus stronger Pb-Br bonding (*38*).

We speculate that the CDB technique is a universal approach to enhance the efficiency and stability of all types of perovskite solar modules. To facilitate mass production in the future, development of a green solvent system (avoiding the use of toxic DMF) should be considered for the manufacturing of all-perovskite tandem solar modules (*39*, *40*).

REFERENCES AND NOTES

1. G. E. Eperon *et al*., *Science* **354**, 861–865 (2016).
2. K. Xiao *et al*., *Nat. Energy* **5**, 870–880 (2020).
3. R. Lin *et al*., *Nat. Energy* **4**, 864–873 (2019).
4. R. Lin *et al*., *Nature* **603**, 73–78 (2022).
5. M. A. Green *et al*., *Prog. Photovolt. Res. Appl.* **30**, 3–12 (2022).
6. J. E. Bishop, J. A. Smith, D. G. Lidzey, *ACS Appl. Mater. Interfaces* **12**, 48237–48245 (2020).
7. X. Peng *et al*., *Adv. Funct. Mater.* **27**, 1703704 (2017).
8. Y. Deng *et al*., *Nat. Energy* **3**, 560–566 (2018).
9. M. Yang *et al*., *Nat. Energy* **2**, 17038 (2017).
10. Z. Yang *et al*., *Sci. Adv.* **7**, eabg3749 (2021).
11. J. B. Whitaker *et al*., *Sustain. Energy Fuels* **2**, 2442–2449 (2018).
12. J. Li *et al*., *Joule* **4**, 1035–1053 (2020).
13. N.-G. Park, *CrystEngComm* **18**, 5977–5985 (2016).
14. Y. Rong *et al*., *Science* **361**, eaat8235 (2018).
15. Y. Hu *et al*., *Joule* **3**, 2076–2085 (2019).
16. E. Bi *et al*., *Nat. Commun.* **8**, 15330 (2017).
17. J. Li, Q. Dong, N. Li, L. Wang, *Adv. Energy Mater.* **7**, 1602922 (2017).
18. H. Gao *et al*., *Sol. RRL* **5**, 2100814 (2021).
19. C. C. Boyd *et al*., *ACS Energy Lett.* **3**, 1772–1778 (2018).
20. Y. Deng *et al*., *Nat. Energy* **6**, 633–641 (2021).
21. Y. Deng *et al*., *Sci. Adv.* **5**, eaax7537 (2019).
22. S. Chen *et al*., *Science* **373**, 902–907 (2021).
23. H. Chen *et al*., *Nature* **550**, 92–95 (2017).
24. N. G. Park, K. Zhu, *Nat. Rev. Mater.* **5**, 333–350 (2020).
25. Z. Liu *et al*., *Nat. Energy* **5**, 596–604 (2020).
26. E. Bi *et al*., *Joule* **3**, 2748–2760 (2019).
27. A. F. Palmstrom *et al*., *Joule* **3**, 2193–2204 (2019).
28. D. Zhao *et al*., *Nat. Energy* **3**, 1093–1100 (2018).
29. Z. Yang *et al*., *Nat. Commun.* **10**, 4498 (2019).
30. J. W. Yoo *et al*., *Joule* **5**, 2420–2436 (2021).
31. S. Jariwala *et al*., *Joule* **3**, 3048–3060 (2019).
32. Y. Zhao *et al*., *Nat. Commun.* **11**, 6328 (2020).
33. R. E. Beal *et al*., *Matter* **2**, 207–219 (2020).
34. K. A. Bush *et al*., *ACS Energy Lett.* **3**, 428–435 (2018).
35. L. A. Castriotta *et al*., *Adv. Energy Mater.* **12**, 2103420 (2022).
36. S. H. Reddy, F. Di Giacomo, A. Di Carlo, *Adv. Energy Mater.* **12**, 2103534 (2022).
37. R. Prasanna *et al*., *Nat. Energy* **4**, 939–947 (2019).
38. J. Yang *et al*., *Nano Energy* **54**, 218–226 (2018).
39. L. Vesce *et al*., *Sol. RRL* **5**, 2100073 (2021).
40. R. Swartwout *et al*., *Sol. RRL* **6**, 2100567 (2021).

ACKNOWLEDGMENTS

Funding: This work was financially supported by the National Natural Science Foundation of China (61974063, 61921005, and U21A2076), the Natural Science Foundation of Jiangsu Province (BK20202008 and BK20190315), Fundamental Research Funds for the Central Universities (0213/14380206 and 0205/14380252), the Frontiers Science Center for Critical Earth Material Cycling Fund (DLTD2109), the Program for Innovative Talents and Entrepreneurs in Jiangsu, and the Engineering and Physical Science Research Council, UK (EP/S004947/1). Y.H. acknowledges support from the National University of Singapore (NUS) Presidential Young Professorship (R-279-000-617-133 and R-279-001-617-133). This work was authored in part by SERIS, a research institute at NUS. SERIS is supported by NUS, the National Research Foundation Singapore (NRF), the Energy Market Authority of Singapore (EMA), and the Singapore Economic Development Board (EDB). R.D.J.O. thanks the Penrose Scholarship for funding his studentship. **Author contributions:** H.T. conceived and directed the overall project. H.T. and H.J.S. supervised the research. K.X. fabricated all the devices and conducted the characterization. M.Z. fabricated NBG perovskite films, R.D.J.O. and Y.-H.L. carried out PLQY measurements, Z.L. carried out grazing-incidence wide-angle x-ray scattering characterizations, X.L. helped with module encapsulation and carried out transient photovoltage measurements, H.L. and R.L. carried out SEM characterizations, and J.X. helped with the ultraviolet-visible absorption and SEM with energy-dispersive x-ray characterizations. Y.H., X.W., and J.L. assisted with data analyses. K.X., Y.-H.L., D.L., Y.H., H.J.S., and H.T. wrote the manuscript. All authors read and commented on the manuscript. **Competing interests:** H.T. and K.X. are inventors on a patent application related to this work filed by Nanjing University. H.J.S. is a co-founder, Chief Scientific Officer, and director of Oxford PV Ltd., a company that is commercializing perovskite PVs. The other authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn7696
Materials and Methods
Figs. S1 to 38
Tables S1 to S6
References (*41*–*47*)
Movies S1 and S2

Submitted 23 December 2021; accepted 5 April 2022
10.1126/science.abn7696

23andMe has

Open Postdoc Positions - Hiring Now!

Interested in joining our diverse team of ever-curious scientists, researchers, and many more who share one common goal - to make a difference in people's lives?

Join our team!

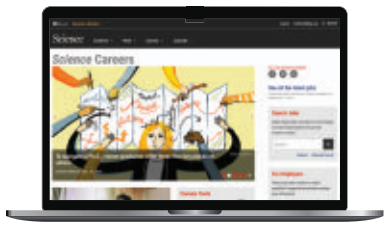
Visit our website for more information: https://research.23andme.com/postdocs/

23andMe

Science Careers helps you advance your career. Learn how!

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.
- Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
- Complete an interactive, personalized career plan at "my IDP."
- Visit our Employer Profiles to learn more about prospective employers.
- Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free



SCIENCECAREERS.ORG

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

HARVARD UNIVERSITY

MCB
Department of Molecular and Cellular Biology

Postdoctoral Research Positions

The Department is actively recruiting postdoctoral fellows interested in a broad range of topics, including cell biology, systems biology, biophysics, biochemistry, developmental biology, neurobiology, genetics, microbiology, computational biology, and structural biology.

We are seeking highly creative and motivated postdocs with broad interests and outstanding graduate training.

We encourage applicants from various backgrounds and promote diversity throughout the University.

Postdoctoral appointments are initially 1 year in duration, renewable on an annual basis.

Location

Cambridge, MA USA

Primary Contact Details

Department of Molecular and Cellular Biology
www.mcb.harvard.edu

Recent Faculty in the Department

Nicholas Bellono

Postdoc openings in the Bellono lab:
Explore molecular adaptations underlying unique behaviors. Octopus, sharks, jellyfish, photosynthetic animals, walking fish, and more! Background in cell biology, physiology, biochemistry, structure, other molecular approaches.

Doeke Hekstra

The Hekstra lab aims to understand how proteins work as evolving, physical "machines". To do so, we develop crystallographic, computational and other methods. Let us know what you are interested in. Experience with computational methods, statistics, physics, and (bio)chemistry is a plus.

Maxim Prigozhin

The Prigozhin lab is looking for postdoctoral fellows to be involved in collaborative and interdisciplinary work in the area of single-cell and single-molecule biophysics of transmembrane signaling. We are particularly interested in cryo-vitrification, all kinds of electron and optical microscopy, and G-protein-coupled receptors. Let us know what you are excited about!

Amanda Whipple

The Whipple lab is investigating mechanisms of gene regulation in genomic imprinting, functions of imprinted non-coding RNAs in the brain, and therapeutic approaches for imprinting disorders. We seek collaborative and cross-disciplinary postdoctoral candidates in the areas of epigenetics, genomics, and neurobiology.

For a Full List of Opportunities and Information on How to Apply, Please Visit:
https://www.mcb.harvard.edu/job-opportunities/post-doctoral-positions/

About Us

Our Science

We research and teach how the collective behavior of molecules and cells forms the basis of life. We are driven by a passion for discovery and value collaborative approaches to scientific inquiry, where connections between people fuel interdisciplinary science and break boundaries across varied experimental systems. Against a backdrop of cutting-edge biological research, we work as a team of educators and mentors to inspire and train the next generation of scientists and global citizens.

Our Community

At the core of the MCB department is a commitment to foster an environment in which all individuals have the opportunity to thrive. It is our shared responsibility to create an inclusive culture, where we support and respect each other as colleagues. We embrace a diverse range of perspectives, expertise, identities, experiences, talents, and abilities. By continually strengthening this foundation of investing in the well-being of our people, we enable our community's growth and pursuit of the creative and innovative approaches that underlie scientific excellence.

Harvard University is an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, sex, gender identity, sexual orientation, religion, creed, national origin, ancestry, age, protected veteran status, disability, genetic information, military service, pregnancy and pregnancy-related conditions, or other protected status.

ScienceCareers online @sciencecareers.org
JOB FOCUS: POSTDOC

By **Yang Xu**

Second chances

When I started graduate school, everything looked bright. As an undergraduate in China, I felt lucky and honored that a South Korean professor visiting my university offered me a place in his lab for a Ph.D. It seemed like an amazing opportunity to get international research experience. But soon things started to go downhill. The stipend was just enough to cover food, rent, and some necessary bills. Even minor expenses such as paying for clothing and social events required me to cut back elsewhere. When I looked into external fellowships, which helped many of my Korean classmates cover their living expenses, I was told that my international training wasn't elite enough to qualify. Despite feeling overwhelmed and isolated, I carried on—trapped by my fear of failure.

Time I could have spent studying I instead dedicated to finding the least expensive groceries I could, and cooking to save money. I relied on rice and bread to fill my stomach, but I often still felt hungry. I was lonely and madly wanted pleasure to escape from the stress. So, I did something I immediately regretted: I had unprotected sex with a man from a dating app. For weeks afterward, I was paranoid I might have contracted a sexually transmitted infection. To make matters worse, I didn't have anyone to turn to for support because I hadn't told anyone I'm gay. I worried that being open about my sexuality would bias my co-workers against me, so it seemed safer to stay quiet.

"Instead of seeing failure as shameful, I could treat it as an opportunity to improve."

Eventually I went to the doctor for a physical. To my relief, my HIV test was negative. But the results weren't all good news. Many of my metabolite levels were abnormal, likely because of my high-carb diet. I also had high cholesterol, possibly because of stress. It was becoming clear that my situation was not healthy or sustainable.

But I had been taught I was not allowed to fail. I feared that leaving my Ph.D. program would be the beginning of a freefall from which I might not recover. Opportunities to work internationally are rare, and I wasn't sure I would find one again. I was also hesitant to talk with my adviser. I didn't want to throw my life problems at him, and I worried I would be letting him down.

Soon, though, I reached my breaking point. I went to my adviser and told him I couldn't continue. To my surprise, he wasn't disappointed in me; instead he said he felt sorry for what I was going through and regretted he hadn't stepped in earlier. I realized one more mistake I made was keeping my problems to myself.

That was 7 years ago. I'm now close to finishing my Ph.D., and I've learned that fear of failure needn't dominate or define me. After returning to China and working as a research assistant, I applied to Ph.D. programs in the United States, where I hoped I would find a better fit. The director of my U.S. program was concerned I might drop out again, but I convinced him the first program was not right for me and I still had the passion to do research. Since then I have had a great journey of learning and growing in an environment that encourages differing viewpoints. I feel valued, respected, and comfortable being open about my sexuality.

It hasn't always been smooth sailing. When I failed my first attempt at the qualifying exam, all my negative feelings around my first Ph.D. experience came flooding back. But instead of trying to ignore them or having too many drinks and hoping they would go away, as I had done in the past, I faced them. I realized the notion that one should never fail is toxic and only creates fear. Instead of seeing failure as shameful, I could treat it as an opportunity to improve. I watched talks on YouTube to learn how to give better presentations. I practiced my writing by imagining an impatient reader whom I can only persuade with clear, concise language. I turned to my colleagues for help. The second time around, I passed my exam.

I am grateful to my program director, who gave me a second chance. And I am proud I managed to leave the graduate school that was not right for me and kept the faith that I would rise again. ■

Yang Xu is a Ph.D. student at the University of Tennessee, Knoxville.
Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

ARE YOU THE KIND WHO WORKS TO HELP ALL HUMANKIND?

Find your next job at ScienceCareers.org

The relevance of science is at an all-time high these days. For anyone who's looking to get ahead in —or just plain get into— science, there's no better, more trusted resource or authority on the subject than *Science* Careers. Here you'll find opportunities and savvy advice across all disciplines and levels. There's no shortage of global problems today that science can't solve. Be part of the solution.

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

You're only an essay away!

Check out our Awards and Prizes that are currently seeking submissions.

Neurobiology
Submissions Opened: February 1
Submissions Close: June 15

Cell and Molecular Biology
Genomics, Proteomics, and Systems Biology
Ecology and Environment
Molecular Medicine
Submissions Opened: March 1
Submissions Close: July 15

Immunology
Submissions Open: May 1
Submissions Close: October 1

Scientific Innovation & Entrepreneurship
Submissions Open: June 1
Submissions Close: November 1

Science Awards and Prizes seeks to find the next generation of early-career researchers on the cusp of changing the scientific landscape. Start your journey by finding the Award or Prize that fits you and find out how to apply today!

Science
AAAS

bit.ly/3qKxMN0

Scan the code to start your application process!